

- Higher eukaryotes transcription initiation factor TFIID 250 Kd subunit (TBP-associated factor p250) (gene CCG1). P250 associated with the TFIID TATA-box binding protein and seems essential for progression of the G1 phase of the cell cycle.
- 5 - Human RING3, a protein of unknown function encoded in the MHC class II locus.
- Mammalian CREB-binding protein (CBP), which mediates cAMP-gene regulation by binding specifically to phosphorylated CREB protein.
- Drosophila female sterile homeotic protein (gene fsh), required maternally for proper expression of other homeotic genes involved in pattern formation, such as Ubx.
- 10 - Drosophila brahma protein (gene brm), a protein required for the activation of multiple homeotic genes.
- Mammalian homologs of brahma. In human, three brahma-like proteins are known: SNF2a(hBRM), SNF2b, and BRG1.
- Human BS69, a protein that binds to adenovirus E1A and inhibits E1A transactivation
- 15 - Human peregrin (or Br140).
- Yeast BDF1 [3], a transcription factor involved in the expression of a broad class of genes including snRNAs.
- Yeast GCN5, a general transcriptional activator operating in concert with certain other DNA-binding transcriptional activators, such as GCN4, HAP2/3/4 or ADA2.
- 20 - Yeast NPS1/STH1, involved in G(2) phase control in mitosis.
- Yeast SNF2/SWI2, which is part of a complex with the SNF5, SNF6, SWI3 and ADR6/SWI1 proteins. This SWI-complex is involved in transcriptional activation.
- Yeast SPT7, a transcriptional activator of Ty elements and possibly other genes.
- Caenorhabditis elegans protein cbp-1.
- 25 - Yeast hypothetical protein YGR056w.
- Yeast hypothetical protein YKR008w.
- Yeast hypothetical protein L9638.1.

30 Some proteins contain a region which, while similar to some extent to a classical bromodomain, diverges from it by either lacking part of the domain or because of an insertion. These proteins are:

- Mammalian protein HRX (also known as All-1 or MLL), a protein involved in translocations leading to acute leukemias and which possibly acts as a transcriptional regulatory factor. HRX contains a region similar to the C-terminal half of the bromodomain.
- *Caenorhabditis elegans* hypothetical protein ZK783.4. The bromodomain of this protein has a 23 amino-acid insertion.
- Yeast protein YTA7. This protein contains a region with significant similarity to the C-terminal half of the bromodomain. As it is a member of the AAA family (see <PDOC00572>) it is also in a functionally different context.

10 The above proteins generally contain a single bromodomain, but some of them contain two copies, this is the case of BDF1, CCG1, fsh, RING3, YKR008w and L9638.1.

15 The exact function of this domain is not yet known but it is thought to be involved in protein-protein interactions and it may be important for the assembly or activity of multicomponent complexes involved in transcriptional activation.

The consensus pattern that has been developed spans a major part of the bromodomain; a more sensitive detection is available through the use of a profile which spans the whole domain.

20 Consensus pattern[STANVF]-x(2)-F-x(4)-[DNS]-x(5,7)-[DENGTF]-Y-[HFY]-x(2)-[LIVMFY]-x(3)-[LIVM]-x(4)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(2)-N-[SACF]-x(2)-[FY]

25 References

- [1] Haynes S.R., Doolard C., Winston F., Beck S., Trowsdale J., Dawid I.B. Nucleic Acids Res. 20:2693-2603(1992).
- [2] Tamkun J.W., Deuring R., Scott M.P., Kissinger M., Pattatucci A.M., Kaufman T.C., Kennison J.A. Cell 68:561-572(1992).
- [3] Tamkun J.W. Curr. Opin. Genet. Dev. 5:473-477(1995).

808. (CH) Actinin-type actin-binding domain signatures

PROSITE cross-reference(s): PS00019; ACTININ_1, PS00020; ACTININ_2

Alpha-actinin is a F-actin cross-linking protein which is thought to anchor actin to a variety of intracellular structures [1]. The actin-binding domain of alpha-actinin seems to reside in the first 250 residues of the protein. A similar actin-binding domain has been found in the N-terminal region of many different actin-binding proteins [2,3]:

- In the beta chain of spectrin (or fodrin).
- In dystrophin, the protein defective in Duchenne muscular dystrophy (DMD) and which may play a role in anchoring the cytoskeleton to the plasma membrane.
- In the slime mold gelation factor (or ABP-120).
- In actin-binding protein ABP-280 (or filamin), a protein that link actin filaments to membrane glycoproteins.
- In fimbrin (or plastin), an actin-bundling protein. Fimbrin differs from the above proteins in that it contains two tandem copies of the actin-binding domain and that these copies are located in the C-terminal part of the protein.

Two conserved regions were selected as signature patterns for this type of main. The first of this region is located at the beginning of the domain, while the second one is located in the central section and has been shown to be essential for the binding of actin.

Consensus pattern[EQ]-x(2)-[ATV]-[FY]-x(2)-W-x-N

Consensus pattern[LIVM]-x-[SGN]-[LIVM]-[DAGHE]-[SAG]-x-[DNEAG]-[LIVM]-x-[DEAG]-x(4)-[LIVM]-x-[LM]-[SAG]-[LIVM]-[LIVMT]-W-x- [LIVM](2)

[1] Schleicher M., Andre E., Harmann A., Noegel A.A. Dev. Genet. 9:521-530(1988).

[2] Matsudaira P. Trends Biochem. Sci. 16:87-92(1991).

[3] Dubreuil R.R. BioEssays 13:219-226(1991).

809. (COX1) Heme-copper oxidase subunit I, copper B binding region signature

PROSITE cross-reference(s): PS00077; COX1

Heme-copper respiratory oxidases [1] are oligomeric integral membrane protein complexes that catalyze the terminal step in the respiratory chain: they transfer electrons from cytochrome c or a quinol to oxygen. Some terminal oxidases generate a transmembrane proton gradient across the plasma membrane (prokaryotes) or the mitochondrial inner membrane (eukaryotes). The enzyme

complex consists of 3-4 subunits (prokaryotes) up to 13 polypeptides (mammals) of which only the catalytic subunit (equivalent to mammalian subunit 1 (CO I)) is found in all heme-copper respiratory oxidases. The presence of a bimetallic center (formed by a high-spin heme and copper B) as well as a low-spin heme, both ligated to six conserved histidine residues near the outer side of four transmembrane spans within CO I is common to all family members [2-4].

In contrary to eukaryotes the respiratory chain of prokaryotes is branched to multiple terminal oxidases. The enzyme complexes vary in heme and copper composition, substrate type and substrate affinity. The different respiratory oxidases allow the cells to customize their respiratory systems according a variety of environmental growth conditions [1].

Recently also a component of an anaerobic respiratory chain has been found to contain the copper B binding signature of this family: nitric oxide reductase (NOR) exists in denitrifying species of Archae and Eubacteria.

Enzymes that belong to this family are:

- Mitochondrial-type cytochrome c oxidase (EC 1.9.3.1) which uses cytochrome c as electron donor. The electrons are transferred via copper A (Cu(A)) and heme a to the bimetallic center of CO I that is formed by a penta-coordinated heme a and copper B (Cu(B)). Subunit 1 contains 12 transmembrane regions. Cu(B) is said to be ligated to three of the conserved histidine residues within the transmembrane segments 6 and 7.
- Quinol oxidase from prokaryotes that transfers electrons from a quinol to the binuclear center of polypeptide I. This category of enzymes includes Escherichia coli cytochrome O terminal oxidase complex which is a component of the aerobic respiratory chain that predominates when cells are grown at high aeration.
- FixN, the catalytic subunit of a cytochrome c oxidase expressed in nitrogen-fixing bacteroids living in root nodules. The high affinity for oxygen allows oxidative phosphorylation under low oxygen concentrations. A similar enzyme has been found in other purple bacteria.

- Nitric oxide reductase (EC 1.7.99.7) from *Pseudomonas stutzeri*. NOR reduces nitrate to dinitrogen. It is a heterodimer of norC and the catalytic subunit norB. The latter contains the 6 invariant histidine residues and 12 transmembrane segments [5].

5

As a signature pattern the copper-binding region was used.

Consensus pattern[YWG]-[LIVFYWTA](2)-[VGS]-H-[LNP]-x-V-x(44,47)-H-H [The three H's are copper B ligands]

10

Notcytochrome bd complexes do not belong to this family.

[1]

Garcia-Horsman J.A., Barquera B., Rumbley J., Ma J., Gennis R.B.
J. Bacteriol. 176:5587-5600(1994).

15

[2]

Castresana J., Luebben M., Saraste M., Higgins D.G.
EMBO J. 13:2516-2525(1994).

[3]

Capaldi R.A., Malatesta F., Darley-Usmar V.M.
Biochim. Biophys. Acta 726:135-148(1983).

20

[4]

Holm L., Saraste M., Wikstrom M.
EMBO J. 6:2819-2823(1987).

25

[5]

Saraste M., Castresana J.
FEBS Lett. 341:1-4(1994).

810. (dehydrog_molyb) Eukaryotic molybdopterin oxidoreductases signature
PROSITE cross-reference(s): PS00559; MOLYBDOPTERIN_EUK

30

A number of different eukaryotic oxidoreductases that require and bind a molybdopterin cofactor have been shown [1] to share a few regions of sequence similarity. These enzymes are:

- Xanthine dehydrogenase (EC 1.1.1.204), which catalyzes the oxidation of xanthine to uric acid with the concomitant reduction of NAD. Structurally, this enzyme of about 1300 amino acids consists of at least three distinct domains: an N-terminal 2Fe-2S ferredoxin-like iron-sulfur binding domain (see <PDOC00175>), a central FAD/NAD-binding domain and a C-terminal Mo-pterin domain.

- Aldehyde oxidase (EC 1.2.3.1), which catalyzes the oxidation aldehydes into acids. Aldehyde oxidase is highly similar to xanthine dehydrogenase in its sequence and domain structure.

- Nitrate reductase (EC 1.6.6.1), which catalyzes the reduction of nitrate to nitrite. Structurally, this enzyme of about 900 amino acids consists of an N-terminal Mo-pterin domain, a central cytochrome b5-type heme-binding domain (see <PDOC00170>) and a C-terminal FAD/NAD-binding cytochrome reductase domain.

- Sulfite oxidase (EC 1.8.3.1), which catalyzes the oxidation of sulfite to sulfate. Structurally, this enzyme of about 460 amino acids consists of an N-terminal cytochrome b5-binding domain followed by a Mo-pterin domain.

There are a few conserved regions in the sequence of the molybdopterin-binding domain of these enzymes. The pattern used to detect these proteins is based on one of them. It contains a cysteine residue which could be involved in binding the molybdopterin cofactor.

Consensus pattern[GA]-x(3)-[KRNQHT]-x(11,14)-[LIVMFYWS]-x(8)-[LIVMF]-x-C-x(2)-[DEN]-R-x(2)-[DE]

[1]

Wootton J.C., Nicolson R.E., Cock J.M., Walters D.E., Burke J.F., Doyle W.A., Bray R.C.
Biochim. Biophys. Acta 1057:157-185(1991).

811. (DNA_ligase) ATP-dependent DNA ligase signatures

PROSITE cross-reference(s): PS00697; DNA_LIGASE_A1, PS00333; DNA_LIGASE_A2

DNA ligase (polydeoxyribonucleotide synthase) is the enzyme that joins two DNA fragments by catalyzing the formation of an internucleotide ester bond between phosphate and deoxyribose. It is active during DNA replication, DNA repair and DNA recombination. There are two forms of DNA ligase: one requires ATP (EC 6.5.1.1), the other NAD (EC 6.5.1.2).

Eukaryotic, archaebacterial, virus and phage DNA ligases are ATP-dependent. During the first step of the joining reaction, the ligase interacts with ATP to form a covalent enzyme-adenylate intermediate. A conserved lysine residue is the site of adenylation [1,2].

Apart from the active site region, the only conserved region common to all ATP-dependent DNA ligases is found [3] in the C-terminal section and contains a conserved glutamate as well as four positions with conserved basic residues.

Signature patterns were developed for both conserved regions.

Consensus pattern[EDQH]-x-K-x-[DN]-G-x-R-[GACIVM] [K is the active site residue]

Consensus patternE-G-[LIVMA]-[LIVM](2)-[KR]-x(5,8)-[YW]-[QNEK]-x(2,6)-[KRH]-x(3,5)-K-[LIVMFY]-K

Sequences known to belong to this class detected by the patternALL, except for archaebacterial DNA ligases.

[1]

Tomkinson A.E., Totty N.F., Ginsburg M., Lindahl T.
Proc. Natl. Acad. Sci. U.S.A. 88:400-404(1991).

[2]

Lindahl T., Barnes D.E.
Annu. Rev. Biochem. 61:251-281(1992).

[3]

Kletzin A.

Nucleic Acids Res. 20:5389-5396(1992).

812. (FAD_Gly3P_dh) FAD-dependent glycerol-3-phosphate dehydrogenase signatures
PROSITE cross-reference(s): PS00977; FAD_G3PDH_1, PS00978; FAD_G3PDH_2

5

FAD-dependent glycerol-3-phosphate dehydrogenase (EC 1.1.99.5) (GPD) catalyzes
the conversion of glycerol-3-phosphate into dihydroxyacetone phosphate. In
bacteria [1] it is associated with the utilization of glycerol coupled to
respiration. In *Escherichia coli*, two isozymes are known: one expressed under
10 anaerobic conditions (gene *glpA*) and one in aerobic conditions (gene *glpD*). In
eukaryotes, a mitochondrial form of GPD participates in the glycerol phosphate
shuttle in conjunction with an NAD-dependent cytoplasmic GPD (EC 1.1.1.8) [2,
3].

15

These enzymes are proteins of about 60 to 70 Kd which contain a probable
FAD-binding domain in their N-terminal extremity. The mammalian enzyme differs
from the bacterial or yeast proteins by having an EF-hand calcium-binding
region (See <PDOC00018>) in its C-terminal extremity.

20

Two signature patterns were developed. One based on the first half of the FAD-
binding domain and one which corresponds to a conserved region in the central
part of these enzymes.

Consensus pattern[IV]-G-G-G-x(2)-G-[STACV]-G-x-A-x-D-x(3)-R-G

25

Consensus patternG-G-K-x(2)-[GSTE]-Y-R-x(2)-A

[1]

Austin D., Larson T.J.

J. Bacteriol. 173:101-107(1991).

30

[2]

Roennow B., Kielland-Brandt M.C.

Yeast 9:1121-1130(1993).

[3]

Brown L.J., McDonald M.J., Lehn D.A., Moran S.M.

J. Biol. Chem. 269:14363-14366(1994).

813. (Fapy_DNA_glyco) Formamidopyrimidine-DNA glycosylase signature

PROSITE cross-reference(s): PS01242; FPG

5

Formamidopyrimidine-DNA glycosylase (EC 3.2.2.23) [1] (Fapy-DNA glycosylase) (gene fpg) is a bacterial enzyme involved in DNA repair and which excise oxidized purine bases to release 2,6-diamino-4-hydroxy-5N-methylformamido-pyrimidine (Fapy) and 7,8-dihydro-8-oxoguanine (8-OxoG) residues. In addition to its glycosylase activity, FPG can also nick DNA at apurinic/apyrimidinic sites (AP sites). FPG is a monomeric protein of about 32 Kd which binds and require zinc for its activity.

10

The binding site for zinc seems to be located in the C-terminal part of the enzyme where four conserved and essential [2] cysteines are located. A signature pattern was developed based on this region.

15

Consensus pattern C-x(2,4)-C-x-[GTAQ]-x-[IV]-x(7)-R-[GSTAN]-[STA]-x-[FYI]-C-x(2)-C-Q

20

[The four C's are putative zinc ligands]

[1]

Duwat P., de Oliveira R., Ehrlich S.D., Boiteux S.

Microbiology 141:411-417(1995).

25

[2]

O'Connor T.E., Graves R.J., Demurcia G., Castaing B., Laval J.

J. Biol. Chem. 268:9063-9070(1993).

814. (G_glu_transpept) Gamma-glutamyltranspeptidase signature

30

PROSITE cross-reference(s): PS00462; G_GLU_TRANSPEPTIDASE

Gamma-glutamyltranspeptidase (EC 2.3.2.2) (GGT) [1] catalyzes the transfer of the gamma-glutamyl moiety of glutathione to an acceptor that may be an amino acid, a peptide or water (forming glutamate). GGT plays a key role in the

gamma-glutamyl cycle, a pathway for the synthesis and degradation of glutathione. In prokaryotes and eukaryotes, it is an enzyme that consists of two polypeptide chains, a heavy and a light subunit, processed from a single chain precursor. The active site of GGT is known to be located in the light subunit.

The sequences of mammalian and bacterial GGT show a number of regions of high similarity [2]. Pseudomonas cephalosporin acylases (EC 3.5.1.-) that convert 7-beta-(4-carboxybutanamido)-cephalosporanic acid (GL-7ACA) into 7-aminocephalosporanic acid (7ACA) and glutaric acid are evolutionary related to GGT and also show some GGT activity [3]. Like GGT, these GL-7ACA acylases, are also composed of two subunits.

One of the conserved regions correspond to the N-terminal extremity of the mature light chains of these enzymes. This region was used as a signature pattern.

Consensus pattern T-[STA]-H-x-[ST]-[LIVMA]-x(4)-G-[SN]-x-V-[STA]-x-T-x-T-[LIVM]-[NE]-x(1,2)-[FY]-G

[1]

Tate S.S., Meister A.

Meth. Enzymol. 113:400-419(1985).

[2]

Suzuki H., Kumagai H., Echigo T., Tochikura T.

J. Bacteriol. 171:5169-5172(1989).

[3]

Ishiye M., Niwa M.

Biochim. Biophys. Acta 1132:233-239(1992).

815. G-protein gamma subunit profile

PROSITE cross-reference(s): PS50058; G_PROTEIN_GAMMA

Guanine nucleotide-binding proteins (G proteins) [1] act as intermediaries in

5

10

The *Caenorhabditis elegans* protein egl-10, which is a regulator of G-protein signalling, contains a G-protein gamma-like domain.

15

[1]

Pennington S.R.

20

816. GNS1/SUR4 family signature

PROSITE cross-reference(s): PS01188; GNS1_SUR4

25

- 30

The proteins have from 290 to 435 amino acid residues. Structurally, they seem to be formed of three sections: a N-terminal region with two transmembrane domains, a central hydrophilic loop and a C-terminal region that contains from one to three transmembrane domains. A conserved region that contains three histidines was selected as a signature pattern. This region is located in the hydrophilic loop.

Consensus pattern L-x-F-L-H-x-Y-H-H

[1]

Bairoch A.

Unpublished observations (1996).

[2]

El-Sherbeini M., Clemas J.A.

J. Bacteriol. 177:3227-3234(1995).

[3]

Garcia-Arranz M., Maldonado A.M., Mazon M.J., Portillo F.

J. Biol. Chem. 269:18076-18082(1994).

817. Immunoglobulins and major histocompatibility complex proteins signature
PROSITE cross-reference(s): PS00290; IG_MHC

The basic structure of immunoglobulin (Ig) [1] molecules is a tetramer of two light chains and two heavy chains linked by disulfide bonds. There are two types of light chains: kappa and lambda, each composed of a constant domain (CL) and a variable domain (VL). There are five types of heavy chains: alpha, delta, epsilon, gamma and mu, all consisting of a variable domain (VH) and three (in alpha, delta and gamma) or four (in epsilon and mu) constant domains (CH1 to CH4).

The major histocompatibility complex (MHC) molecules are made of two chains. In class I [2] the alpha chain is composed of three extracellular domains, a transmembrane region and a cytoplasmic tail. The beta chain (beta-2-

microglobulin) is composed of a single extracellular domain. In class II [3], both the alpha and the beta chains are composed of two extracellular domains, a transmembrane region and a cytoplasmic tail.

It is known [4,5] that the Ig constant chain domains and a single extracellular domain in each type of MHC chains are related. These homologous domains are approximately one hundred amino acids long and include a conserved intradomain disulfide bond. A small pattern around the C-terminal cysteine is involved in this disulfide bond which can be used to detect these category of Ig related proteins.

Consensus pattern[FY]-x-C-x-[VA]-x-H-Sequences known to belong to this class detected by the pattern: Ig heavy chains type Alpha C region : All, in CH2 and CH3. Ig heavy chains type Delta C region : All, in CH3. Ig heavy chains type Epsilon C region: All, in CH1, CH3 and CH4. Ig heavy chains type Gamma C region : All, in CH3 and also CH1 in some cases Ig heavy chains type Mu C region : All, in CH2, CH3 and CH4. Ig light chains type Kappa C region : In all CL except rabbit and Xenopus. Ig light chains type Lambda C region : In all CL except rabbit. MHC class I alpha chains : All, in alpha-3 domains, including in the cytomegalovirus MHC-1 homologous protein [6]. Beta-2-microglobulin : All. MHC class II alpha chains: All, in alpha-2 domains. MHC class II beta chains: All, in beta-2 domains.

[1]

Gough N.
Trends Biochem. Sci. 6:203-205(1981).

[2]

Klein J., Figueroa F.
Immunol. Today 7:41-44(1986).

[3]

Figueroa F., Klein J.
Immunol. Today 7:78-81(1986).

[4]

Orr H.T., Lancet D., Robb R.J., Lopez de Castro J.A., Strominger J.L.

Nature 331:269-272(1988).

096980 = 10100

15

20

- 25

30

Sequences known to belong to this class detected by the patternALL, except for IGFBP-6's.

[1]

Rechler M.M.

Vitam. Horm. 47:1-114(1993).

[2]

Shimasaki S., Ling N.

Prog. Growth Factor Res. 3:243-266(1991).

[3]

Clemmons D.R.

Trends Endocrinol. Metab. 1:412-417(1990).

[4]

Bradham D.M., Igarashi A., Potter R.L., Grotendorst G.R.

J. Cell Biol. 114:1285-1294(1991).

[5]

Maloisel V., Martinerie C., Dambrine G., Plassiart G., Brisac M., Crochet

J., Perbal B.

Mol. Cell. Biol. 12:10-21(1992).

819. LMWPc : Low molecular weight phosphotyrosine protein phosphatase

Number of members: 34

[1]Medline: 94329182, The crystal structure of a low-molecular-weight phosphotyrosine protein phosphatase. Su XD, Taddei N, Stefani M, Ramponi G, Nordlund P; Nature 1994;370:575-578.

820. (myosin_head) ATP/GTP-binding site motif A (P-loop)

PROSITE cross-reference(s): PS00017; ATP_GTP_A

From sequence comparisons and crystallographic data analysis it has been shown [1,2,3,4,5,6] that an appreciable proportion of proteins that bind ATP or GTP share a number of more or less conserved sequence motifs. The best conserved of these motifs is a glycine-rich region, which typically forms a flexible loop between a beta-strand and an alpha-helix. This loop interacts with one of the phosphate groups of the nucleotide. This sequence motif is generally referred to as the 'A' consensus sequence [1] or the 'P-loop' [5].

There are numerous ATP- or GTP-binding proteins in which the P-loop is found. A number of protein families for which the relevance of the presence of such motif has been noted is listed below:

- ATP synthase alpha and beta subunits (see <PDOC00137>).
- Myosin heavy chains.
- Kinesin heavy chains and kinesin-like proteins (see <PDOC00343>).
- Dynamins and dynamin-like proteins (see <PDOC00362>).
- Guanylate kinase (see <PDOC00670>).
- Thymidine kinase (see <PDOC00524>).
- Thymidylate kinase (see <PDOC01034>).
- Shikimate kinase (see <PDOC00868>).
- Nitrogenase iron protein family (nifH/frxC) (see <PDOC00580>).
- ATP-binding proteins involved in 'active transport' (ABC transporters) [7] (see <PDOC00185>).
- DNA and RNA helicases [8,9,10].
- GTP-binding elongation factors (EF-Tu, EF-1alpha, EF-G, EF-2, etc.).
- Ras family of GTP-binding proteins (Ras, Rho, Rab, Ral, Ypt1, SEC4, etc.).
- Nuclear protein ran (see <PDOC00859>).
- ADP-ribosylation factors family (see <PDOC00781>).
- Bacterial dnaA protein (see <PDOC00771>).
- Bacterial recA protein (see <PDOC00131>).
- Bacterial recF protein (see <PDOC00539>).
- Guanine nucleotide-binding proteins alpha subunits (Gi, Gs, Gt, G0, etc.).
- DNA mismatch repair proteins mutS family (See <PDOC00388>).
- Bacterial type II secretion system protein E (see <PDOC00567>).

Not all ATP- or GTP-binding proteins are picked-up by this motif. A number of proteins escape detection because the structure of their ATP-binding site is completely different from that of the P-loop. Examples of such proteins are the E1-E2 ATPases or the glycolytic kinases. In other ATP- or GTP-binding proteins the flexible loop exists in a slightly different form; this is the case for tubulins or protein kinases. A special mention must be reserved for

05063980 101000

5

Walker J.E., Saraste M., Runswick M.J., Gay N.J.
EMBO J. 1:945-951(1982).

10

FEBS Lett. 186:1-7(1985).

Fry D.C., Kuby S.A., Mildvan A.S.

15

Dever T.E., Glynias M.J., Merrick W.C.

[5]

20

[6]

J. Mol. Biol. 229:1165-1174(1993).

25

J. Bioenerg. Biomembr. 22:571-592(1990).

Hodgman T.C.

30

[9]

Nature 337:121-122(1989).

[10]

Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M.

Nucleic Acids Res. 17:4713-4730(1989).

5 821. PE: PE family

This family named after a PE motif near to the amino terminus of the domain. The PE family of proteins all contain an amino-terminal region of about 110 amino acids. The carboxyl terminus of this family are variable and fall into several classes. The largest class of PE proteins is the highly repetitive PGRS class which have a high glycine content. The function of these proteins is uncertain but it has been suggested that they may be related to antigenic variation of Mycobacterium tuberculosis [1]. Number of members: 88

[1] Medline: 98295987. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Barrell BG, et al; Nature 1998;393:537-544.

822. (RNB) Ribonuclease II family signature

PROSITE cross-reference(s): PS01175; RIBONUCLEASE_II

On the basis of sequence similarities, the following bacterial and eukaryotic proteins seem to form a family:

- Escherichia coli and related bacteria ribonuclease II (EC 3.1.13.1) (RNase II) (gene rnb) [1]. RNase II is an exonuclease involved in mRNA decay. It degrades mRNA by hydrolyzing single-stranded polyribonucleotides processively in the 3' to 5' direction.

- Bacterial protein vacB. In Shigella flexneri, vacB has been shown to be required for the expression of virulence genes at the posttranscriptional level.

- Yeast protein SSD1 (or SRK1) which is implicated in the control of the cell cycle G1 phase.

- Yeast protein DIS3 [2], which binds to ran (GSP1) and enhances the

09689980 "101300

nucleotide-releasing activity of RCC1 on ran.

- Fission yeast protein dis3, which is implicated in mitotic control.
- *Neurospora crassa* cyt-4, a mitochondrial protein required for RNA 5' and 3' end processing and splicing.
- 5 - Yeast protein MSU1, which is involved in mitochondrial biogenesis.
- *Synechocystis* strain PCC 6803 protein zam [3], which control resistance to the carbonic anhydrase inhibitor acetazolamide.
- *Caenorhabditis elegans* hypothetical protein F48E8.6.

10 The size of these proteins range from 644 residues (rnb) to 1250 (SSD1). While their sequence is highly divergent they share a conserved domain in their C-terminal section [4]. It is possible that this domain plays a role in a putative exonuclease function that would be common to all these proteins. A signature pattern was developed based on the core of this conserved domain.

15 Consensus pattern[HI]-[FYE]-[GSTAM]-[LIVM]-x(4,5)-Y-[STAL]-x-[FWVAC]-[TV]-[SA]-P-[LIVMA]-[RQ]-[KR]-[FY]-x-D-x(3)-[HQ]

[1]

20 Zilhao R., Camelo L., Arraiano C.M.
Mol. Microbiol. 8:43-51(1993).

[2]

Noguchi E., Hayashi N., Azuma Y., Seki T., Nakamura M., Nakashima N., Yanagida M., He X., Mueller U., Sazer S., Nishimoto T.

25 EMBO J. 15:5595-5605(1996).

[3]

Beuf L., Bedu S., Cami B., Joset F.
Plant Mol. Biol. 27:779-788(1995).

[4]

30 Mian I.S.
Nucleic Acids Res. 25:3187-3195(1997).

823. Src homology 2 (SH2) domain profile

PROSITE cross-reference(s): PS50001; SH2

The Src homology 2 (SH2) domain is a protein domain of about 100 amino-acid residues first identified as a conserved sequence region between the oncoproteins Src and Fps [1]. Similar sequences were later found in many other intracellular signal-transducing proteins [2]. SH2 domains function as regulatory modules of intracellular signalling cascades by interacting with high affinity to phosphotyrosine-containing target peptides in a sequence-specific and strictly phosphorylation-dependent manner [3,4,5,6].

The SH2 domain has a conserved 3D structure consisting of two alpha helices and six to seven beta-strands. The core of the domain is formed by a continuous beta-meander composed of two connected beta-sheets [7].

So far, SH2 domains have been identified in the following proteins:

- Many vertebrate, invertebrate and retroviral cytoplasmic (non-receptor) protein tyrosine kinases. In particular in the Src, Abl, Bkt, Csk and ZAP70 families of kinases.
- Mammalian phosphatidylinositol-specific phospholipase C gamma-1 and -2. Two copies of the SH2 domain are found in those proteins in between the catalytic 'X-' and 'Y-boxes' (see <PDOC50007>).
- Mammalian phosphatidyl inositol 3-kinase regulatory p85 subunit.
- Some vertebrate and invertebrate protein-tyrosine phosphatases.
- Mammalian Ras GTPase-activating protein (GAP).
- Adaptor proteins mediating binding of guanine nucleotide exchange factors to growth factor receptors: vertebrate GRB2, Caenorhabditis elegans sem-5 and Drosophila DRK.
- Mammalian Vav oncoprotein, a guanine-nucleotide exchange factor of the CDC24 family.
- Miscellaneous proteins interacting with vertebrate receptor protein tyrosine kinases: oncoprotein Crk, mammalian cytoplasmic proteins Nck, Shc.
- STAT proteins (signal transducers and activators of transcription).
- Chicken tensin.
- Yeast transcriptional control protein SPT6.

Figure 1 consists of 15 small plots arranged in a grid. Each plot shows the distribution of the number of non-zero elements in the vector x for a specific value of n (from 1 to 15). The x-axis represents the number of non-zero elements, and the y-axis represents the probability or frequency. The distributions are centered around zero non-zero elements, with the spread increasing as n increases.

[1]

10

[2]

[3]

15

[4]

20

[5]

[6]

25

[7]

Kuriyan J., Cowburn D.
Curr. Opin. Struct. Biol. 3:828-837(1993).

824. Sulfate transporters signature

PROSITE cross-reference(s): PS01130; SULFATE TRANSP

30

A number of proteins involved in the transport of sulfate across a membrane as well as some yet uncharacterized proteins have been shown [1,2] to be evolutionary related. These proteins are:

- *Neurospora crassa* sulfate permease II (gene *cys-14*).
- Yeast sulfate permeases (genes *SUL1* and *SUL2*).
- Rat sulfate anion transporter 1 (*SAT-1*).
- Mammalian DTDST, a probable sulfate transporter which, in Human, is involved in the genetic disease, diastrophic dysplasia (DTD).
- Sulfate transporters 1, 2 and 3 from the legume *Stylosanthes hamata*.

- Human pendrin (gene *PDS*), which is involved in a number of hearing loss genetic diseases.

- Human protein DRA (Down-Regulated in Adenoma).
- Soybean early nodulin 70.
- *Escherichia coli* hypothetical protein ychM.
- *Caenorhabditis elegans* hypothetical protein F41D9.5.

As expected by their transport function, these proteins are highly hydrophobic and seem to contain about 12 transmembrane domains. The best conserved region seems to be located in the second transmembrane region and is used as a signature pattern.

Consensus pattern[PAV]-x-Y-[GS]-L-Y-[STAG](2)-x(4)-[LIVFYA]-[LIVST]-[YI]-x(3)-[GA]-[GST]-S-[KR]

[1]

Sandal N.N., Marcker K.A.

Trends Biochem. Sci. 19:19-19(1994).

[2]

Smith F.W., Hawkesford M.J., Prosser I.M., Clarkson D.T.

Mol. Gen. Genet. 247:709-715(1995).

825. TYA: TYA transposon protein

Ty are yeast transposons. A 5.7kb transcript codes for p3 a fusion protein of TYA and TYB. The TYA protein is analogous to the gag protein of retroviruses. TYA a is cleaved to form 46kd protein which can form mature virion like particles [1]. Number of members: 59

[illegible]

5

Class II Aldolase and Adducin N-terminal domain.

-!- This family includes class II aldolases and adducins which have not been ascribed any enzymatic function. Number of members: 37

10

[1] Medline: 93294819. The spatial structure of the class II L-fuculose-1-phosphate aldolase from *Escherichia coli*. Dreyer MK, Schulz GE; J Mol Biol 1993;231:549-553.

15

827. CBD 2

-!- Two tryptophan residues are involved in cellulose binding.

!- Cellulose binding domain found in bacteria. Number of members: 51

20

References:

25

828. P

30

Number of members: 91

References:

[2] Medline: 98225190. Regulatory roles of the P domain of the subtilisin-like prohormone convertases. Zhou A, Martin S, Lipkind G, LaMendola J, Steiner DF; J Biol Chem 1998;273:11107-11114.

PROSITE cross-reference(s): PS01261; UPF0020

- *Escherichia coli* hypothetical protein ycbY and HI0116/15, the corresponding *Haemophilus influenzae* protein.

- *Bacillus subtilis* hypothetical protein ypsC.
- *Synechocystis* strain PCC 6803 hypothetical protein slr0064.
- *Methanococcus jannaschii* hypothetical proteins MJ0438 and MJ0710.

These are hydrophilic proteins of from 40 Kd to about 80 Kd. They can be picked up in the database by the following pattern.

Consensus patternD-P-[LIVMF]-C-G-[ST]-G-x(3)-[LI]-E

References:

[1] Bairoch A. Unpublished observations (1997).

830. Uncharacterized protein family UPF0031 signatures

PROSITE cross-reference(s): PS01049; UPF0031 1; PS01050; UPF0031 2

The following uncharacterized proteins have been shown [1] to share regions of similarities:

- Yeast chromosome XI hypothetical protein YKL151c.
- *Caenorhabditis elegans* hypothetical protein R107.2.
- *Escherichia coli* hypothetical protein yjeF.

Figure 1 consists of 11 bar charts, labeled (a) through (k), arranged vertically. Each chart displays the percentage of total protein in a specific fraction (A, B, C, D, E, F, G, H, I, J, K) for five different treatment groups: Control, 100 mg/kg, 200 mg/kg, 400 mg/kg, and 800 mg/kg. The y-axis for all charts represents the percentage of total protein, ranging from 0 to 100. The x-axis for each chart lists the fractions. The data shows that as the treatment dose increases, the percentage of total protein in most fractions also increases, with the 800 mg/kg group consistently showing the highest percentages across all fractions.

-
- Figure 1 displays 12 histograms showing the distribution of the number of non-zero elements in the sparse matrix A for different values of n (100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200). The x-axis represents the number of non-zero elements, and the y-axis represents the frequency. The distributions are centered around 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 11000, and 12000 respectively. The histograms are arranged in a 4x3 grid.

10

Consensus pattern[GA]-G-x-G-D-[TV]-[LT]-[STA]-G-x-[LIVM]

15

20

25

30

This is a family of bifunctional enzymes catalysing the last steps in de novo purine biosynthesis. The bifunctional enzyme is found in both prokaryotes and eukaryotes. The second last step is catalysed by 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase EC:2.1.2.3 (AICARFT), this enzyme catalyses the formylation of AICAR

with 10-formyl-tetrahydrofolate to yield FAICAR and tetrahydrofolate [1]. The last step is catalysed by IMP (Inosine monophosphate) cyclohydrolase EC:3.5.4.10 (IMPCHase), cyclizing FAICAR (5-formylaminoimidazole-4-carboxamide ribonucleotide) to IMP [1].

5 Number of members: 22

[1] Akira T, Komatsu M, Nango R, Tomooka A, Konaka K, Yamauchi M, Kitamura Y, Nomura S, Tsukamoto I; Medline: 97473523 Molecular cloning and expression of a rat cDNA encoding 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase" [published erratum appears in Gene 1998 Feb 27;208(2):337] Gene 1997;197:289-293.

[2] Rayl EA, Moroson BA, Beardsley GP; Medline: 96147205 The human purH gene product, 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase. Cloning, sequencing, expression, purification, kinetic analysis, and domain mapping." J Biol Chem 1996;271:2225-2233.

833. (AOX)

Alternative oxidase

The alternative oxidase is used as a second terminal oxidase in the mitochondria, electrons are transferred directly from reduced ubiquinol to oxygen forming water [2]. This is not coupled to ATP synthesis and is not inhibited by cyanide, this pathway is a single step process [1]. In rice the transcript levels of the alternative oxidase are increased by low temperature [1].

Number of members: 27

[1] Ito Y, Saisho D, Nakazono M, Tsutsumi N, Hirai A; Medline: 98086211 Transcript levels of tandem-arranged alternative oxidase genes in rice are increased by low temperature." Gene 1997;203:121-129.

5

Protein kinases signatures and profile

10

Eukaryotic protein kinases [1 to 5] are enzymes that belong to a very extensive family of proteins which share a conserved catalytic core common to both serine/threonine and tyrosine protein kinases. There are a number of conserved regions in the catalytic domain of protein kinases. Two of these regions have been selected to build signature patterns. The first region, which is located in the N-terminal extremity of the catalytic domain, is a glycine-rich stretch of residues in the vicinity of a lysine residue, which has been shown to be involved in ATP binding. The second region, which is located in the central part of the catalytic domain, contains a conserved aspartic acid residue which is important for the catalytic activity of the enzyme [6]; two signature patterns were derived for that region: one specific for serine/threonine kinases and the other for tyrosine kinases. A profile was developed which is based on the alignment in [1] and covers the entire catalytic domain.

25

Consensus pattern: [LIV]-G-[P]-G-[P]-[FYWMGSTNH]-[SGA]-{PW}-[LIVCAT]-{PD}-x-[GSTACLIVMFY]-x(5,18)-[LIVMFYWCSTAR]-[AIVP]-[LIVMFAGCKR]-K [K binds ATP]

30

Sequences known to belong to this class detected by the pattern the majority of known protein kinases but it fails to find a number of them, especially viral kinases which are quite divergent in this region and are completely missed by this pattern.

Consensus pattern: [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-K-x(2)-N-[LIVMFYCT](3) [D is an active site residue]

Sequences known to belong to this class detected by the pattern. Most serine/ threonine specific protein kinases with 10 exceptions (half of them viral kinases) and also Epstein-Barr virus BGLF4 and Drosophila ninaC which have respectively Ser and Arg instead of the conserved Lys and which are therefore detected by the tyrosine kinase specific pattern described below.

Consensus pattern: [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-[RSTAC]-x(2)-N-[LIVMFYC](3)
[D is an active site residue] tyrosine specific protein kinases with the exception of human ERBB3 and mouse blk. This pattern will also detect most bacterial aminoglycoside phosphotransferases [8,9] and herpesviruses ganciclovir kinases [10]; which are proteins structurally and evolutionary related to protein kinases. Sequences known to belong to this class detected by the profile ALL, except for three viral kinases. This profile also detects receptor guanylate cyclases (see <PDOC00430>) and 2-5A-dependent ribonucleases.

Sequence similarities between these two families and the eukaryotic protein kinase family have been noticed before. It also detects Arabidopsis thaliana kinase- like protein TMKL1 which seems to have lost its catalytic activity.

Note if a protein analyzed includes the two protein kinase signatures, the probability of it being a protein kinase is close to 100%. Note eukaryotic-type protein kinases have also been found in prokaryotes such as Myxococcus xanthus [11] and Yersinia pseudotuberculosis. Note the patterns shown above has been updated since their publication in [7]. Note this documentation entry is linked to both signature patterns and a profile. As the profile is much more sensitive than the patterns, you should use it if you have access to the necessary software tools to do so.

References

- [1] Hanks S.K., Hunter T., FASEB J. 9:576-596(1995).
- [2] Hunter T., Meth. Enzymol. 200:3-37(1991).
- [3] Hanks S.K., Quinn A.M., Meth. Enzymol. 200:38-62(1991).
- [4] Hanks S.K., Curr. Opin. Struct. Biol. 1:369-383(1991).
- [5] Hanks S.K., Quinn A.M., Hunter T., Science 241:42-52(1988).
- [6] Knighton D.R., Zheng J., Ten Eyck L.F., Ashford V.A., Xuong N.-H., Taylor, S.S., Sowadski J.M., Science 253:407-414(1991).

[7] Bairoch A., Claverie J.-M., Nature 331:22(1988).

[8] Benner S., Nature 329:21-21(1987).

[9] Kirby R., J. Mol. Evol. 30:489-492(1992).

[10] Littler E., Stuart A.D., Chee M.S., Nature 358:160-162(1992).

5 [11] Munoz-Dorado J., Inouye S., Inouye M., Cell 67:995-1006(1991).

835. (Asp_Glu_race)

Aspartate and glutamate racemases signatures

10

Cross-reference(s) PS00923; ASP_GLU_RACEMASE_1 PS00924;
ASP_GLU_RACEMASE_2

Aspartate racemase (EC 5.1.1.13) and glutamate racemase (EC 5.1.1.3) are two evolutionary
15 related bacterial enzymes that do not seem to require a cofactor for their activity [1].

Glutamate racemase, which interconverts L-glutamate into D-glutamate, is required for the
biosynthesis of peptidoglycan and some peptide-based antibiotics such as gramicidin S. In
addition to characterized aspartate and glutamate racemases, this family also includes a
hypothetical protein from Erwinia carotovora and one from Escherichia coli (ygeA). Two
20 conserved cysteines are present in the sequence of these enzymes. They are expected to play
a role in catalytic activity by acting as bases in proton abstraction from the substrate.

Signature patterns were developed for both cysteines.

Consensus pattern: [IVA]-[LIVM]-x-C-x(0,1)-N-[ST]-[MSA]-[STH]-[LIVFYSTANK]

25

Consensus pattern: [LIVM](2)-x-[AG]-C-T-[DEH]-[LIVMFY]-[PNGRS]-x-[LIVM]

[1] Gallo K.A., Knowles J.R., Biochemistry 32:3981-3990(1993).

30

836. (ATP-sulfurylase)

ATP-sulfurylase

09689980-101300

This family consists of ATP-sulfurylase or sulfate adenylyltransferase EC:2.7.7.4 some of which are part of a bifunctional polypeptide chain associated with adenosyl phosphosulphate (APS) kinase APS_kinase. Both enzymes are required for PAPS (phosphoadenosine-phosphosulfate) synthesis from inorganic sulphate [2]. ATP sulfurylase catalyses the synthesis of adenosine-phosphosulfate APS from ATP and inorganic sulphate [1].

Number of members: 37

[1] Kurima K, Warman ML, Krishnan S, Domowicz M, Krueger RC Jr, Deyrup A, Schwartz NB; Medline: 98337975 A member of a family of sulfate-activating enzymes causes murine brachymorphism" [published erratum appears in Proc Natl Acad Sci U S A 1998 Sep 29;95(20):12071] Proc Natl Acad Sci U S A 1998;95:8681-8685.

[2] Rosenthal E, Leustek T; Medline: 96096529 A multifunctional Urechis caupo protein, PAPS synthetase, has both ATP sulfurylase and APS kinase activities." Gene 1995;165:243-248.

837. (ATP-synt_F)

ATP synthase (F/14-kDa) subunit

This family includes 14-kDa subunit from vATPases [1], which is in the peripheral catalytic part of the complex [2]. The family also includes archaebacterial ATP synthase subunit F [3].

Number of members: 23

[1] Guo Y, Kaiser K, Wieczorek H, Dow JA; Medline: 96269411 The Drosophila melanogaster gene vha14 encoding a 14-kDa F-subunit of the vacuolar ATPase." Gene 1996;172:239-243.

[2] Peng SB, Crider BP, Tsai SJ, Xie XS, Stone DK; Medline: 96216416 Identification of a 14-kDa subunit associated with the catalytic sector of clathrin-coated vesicle H⁺-ATPase." J Biol Chem 1996;271:3324-3327.

Subunit structure and organization of the genes of the A1A0 ATPase from the Archaeon *Methanosarcina mazei* Go1." J Biol Chem 1996;271:18843-18852.

Starch binding domain

•

The function of CbiX is uncertain, however it is found in cobalamin biosynthesis operons and so may have a related function. Some CbiX proteins contain a striking histidine-rich region at their C-terminus, which suggests that it might be involved in metal chelation [1].

[1] Raux E, Lanois A, Warren MJ, Rambach A, Thermes C; Medline: 98416126 Cobalamin (vitamin B12) biosynthesis: identification and characterization of a *Bacillus megaterium* cobI operon.” *Biochem J* 1998;335:159-166.

Respiratory-chain NADH dehydrogenase 51 Kd subunit signatures Cross-reference(s)
PS00644; COMPLEX1_51K_1 PS00645; COMPLEX1_51K_2

30 Respiratory-chain NADH dehydrogenase (EC 1.6.5.3) [1,2] (also known as complex I or NADH-ubiquinone oxidoreductase) is an oligomeric enzymatic complex located in the inner mitochondrial membrane which also seems to exist in the chloroplast and in cyanobacteria (as a NADH-plastoquinone oxidoreductase). Among the 25 to 30 polypeptide subunits of this bioenergetic enzyme complex there is one with a molecular weight of 51 Kd (in mammals),

[illegible]

Diaminopimelate epimerase (EC 5.1.1.7) catalyzes the isomerization of L,L- to D,L-meso-diaminopimelate in the biosynthetic pathway leading from aspartate to lysine. This enzyme is a protein of about 30 Kd. Two conserved cysteines seem [1] to function as the acid and base in the catalytic mechanism. As a signature pattern, the region surrounding the first of these two active site cysteines were selected.

Consensus pattern: N-x-D-G-S-x(4)-C-G-N-[GA]-x-R [C is an active site residue] Sequences known to belong to this class detected by the pattern ALL, except for an *Anabaena* dapF which has a Ser instead of the active site Cys.

5

[1] Cirilli M., Zheng R., Scapin G., Blanchard J.S., *Biochemistry* 37:16452-16458(1998).

842. (DNA_gyraseB_C)

10 DNA topoisomerase II signature

Cross-reference(s) PS00177; TOPOISOMERASE_II

DNA topoisomerase I (EC 5.99.1.2) [1,2,3,4,E1] is one of the two types of enzyme that catalyze the interconversion of topological DNA isomers. Type II topoisomerases are ATP-

15

Topoisomerase II is found in phages, archaeobacteria, prokaryotes, eukaryotes, and in African Swine Fever virus (ASF). In bacteriophage T4 topoisomerase II consists of three subunits (the product of genes 39, 52 and 60). In prokaryotes and in archaeobacteria the enzyme, known as DNA gyrase, consists of two subunits (genes *gyrA* and *gyrB* [E2]). In some bacteria, a second type II topoisomerase has been identified; it is known as topoisomerase IV and is required for chromosome segregation, it also consists of two subunits (genes *parC* and *parE*). In eukaryotes, type II topoisomerase is a homodimer.

20

There are many regions of sequence homology between the different subtypes of topoisomerase II. The relation between the different subunits is shown in the following representation:

25

<-----About-1400-residues----->

30

[-----Protein 39-*-----][----Protein 52----] Phage T4

[-----gyrB-----*-----][-----gyrA-----] Prokaryote II

Archaeobacteria

[-----parE-----*-----][-----parD-----] Prokaryote IV

[-----*-----] Eukaryote and

09669980-101300

ASF

'*': Position of the pattern.

5

Consensus pattern: [LIVMA]-x-E-G-[DN]-S-A-x-[STAG]

- 10

15

Protein of unknown function

20

Number of members: 26

- 25

30

Domain of unknown function

This transmembrane region has no known function. Many of the sequences in this family are annotated as hemolysins, however this is due to a similarity to Swiss:Q54318 that does not

[illegible]

contain this domain. This domain is found in the N-terminus of the proteins adjacent to two intracellular CBS domains CBS.

Number of members: 42

5

845. (DUF56)

Integral membrane protein

10

The members of this family are putative integral membrane proteins. The function of the family is unknown, however the family includes Sec59 from yeast. Sec59 is a dolichol kinase EC:2.7.1.108, but it is not clear if the enzymatic activity resides in this region or its N terminal region.

15

Number of members: 13

846. (DUF94)

20

Domain of unknown function

The function of this domain is unknown. It is found in both eukaryotes and archaeobacteria. The alignment contains a completely conserved aspartate residue that may be functionally important. The eukaryotic domains contains three conserved cysteines and a histidine that might be metal binding, however these are absent in the archaeobacterial proteins.

25

Number of members: 9

30

847. (FF)

FF domain

09689980-101300
003T0T" 00689960

This domain may be involved in protein-protein interaction [1].

Number of members: 42

- 5 [1] Bedford MT, Leder P; Medline: 99322199 The FF domain: a novel motif that often
accompanies WW domains.” Trends Biochem Sci 1999;24:264-265.

848. (FLO_LFY)

10 Floricaula / Leafy protein

This family consists of various plant development proteins which are homologues of floricaula (FLO) and Leafy (LFY) proteins which are floral meristem identity proteins. Mutations in the sequences of these proteins affect flower and leaf development.

Number of members: 16

- [1] Hofer J, Turner L, Hellens R, Ambrose M, Matthews P, Michael A, Ellis N; Medline: 97411151 UNIFOLIATA regulates leaf and flower morphogenesis in pea.” *Curr Biol* 1997;7:581-587.

- [2] Weigel D, Alvarez J, Smyth DR, Yanofsky MF, Meyerowitz EM; Medline: 92274452
LEAFY controls floral meristem identity in Arabidopsis." Cell 1992;69:843-859.

849. (G-patch)

G-patch domain

This domain is found in a number of RNA binding proteins, and is also found in proteins that contain RNA binding domains. This suggests that this domain may have an RNA binding function. This domain has seven highly conserved glycines.

Number of members: 47

5

General diffusion Gram-negative porins signature

10

15

- 20

30

Consensus pattern: [LIVMFY]-x(2)-G-x(2)-Y-x-F-x-K-x(2)-[SN]-[STAV]-[LIVMFYW]-V

[1] Benz R., Bauer K., Eur. J. Biochem. 176:1-19(1988).

[2] Jap B.K., Walian P.J., Q. Rev. Biophys. 23:367-403(1990).

[3] Jeanteur D., Lakey J.H., Pattus F., Mol. Microbiol. 5:2153-2164(1991).

5 851. (HlyD)

HlyD family secretion proteins signature

Cross-reference(s) PS00543; HLYD_FAMILY

10 Gram-negative bacteria produce a number of proteins which are secreted into the growth medium by a mechanism that does not require a cleaved N-terminal signal sequence. These proteins, while having different functions, require the help of two or more proteins for their secretion across the cell envelope. Amongst which a protein belonging to the ABC transporters family (see the relevant entry <PDOC00185>) and a protein belonging to a family which is currently composed [1 to 5] of the following members:

15	Gene	Species	Protein which is exported
	-----		-----
	hlyD	Escherichia coli	Hemolysin
	appD	A.pleuropneumoniae	Hemolysin
	lcnD	Lactococcus lactis	Lactococcin A
20	lktD	A.actinomycetemcomitans	Leukotoxin
		Pasteurella haemolytica	
	rtxD	A.pleuropneumoniae	Toxin-III
	cyaD	Bordetella pertussis	Calmodulin-sensitive adenylate cyclase-
			hemolysin (cyclolysin)
25	cvaA	Escherichia coli	Colicin V
	prtE	Erwinia chrysanthemi	Extracellular proteases B and C
	aprE	Pseudomonas aeruginosa	Alkaline protease
	emrA	Escherichia coli	Drugs and toxins
	yjcR	Escherichia coli	Unknown

30 These proteins are evolutionary related and consist of from 390 to 480 amino acid residues. They seem to be anchored in the inner membrane by a N-terminal transmembrane region. Their exact role in the secretion process is not yet known. The C-terminal section of these proteins is the best conserved region; a signature pattern from that region was derived.

09689980 "101300

Consensus pattern: [LIVM]-x(2)-G-[LM]-x(3)-[STGAV]-x-[LIVMT]-x-[LIVMT]-[GE]-x-[KR]-x-[LIVMFYW](2)-x-[LIVMFYW](3)

Sequences known to belong to this class detected by the pattern ALL, except for emrA and yjcR.

5

References:

[1] Gilson L., Mahanty H.K., Kolter R., EMBO J. 9:3875-3884(1990).

[2] Letoffe S., Delepelaire P., Wandersman C., EMBO J. 9:1375-1382(1990).

[3] Stoddard G.W., Petzel J.P., van Belkum M.J., Kok J., McKay L.L., Appl. Environ.

10 Microbiol. 58:1952-1961(1992).

[4] Duong F., Lazdunski A., Cami B., Murgier M., Gene 121:47-54(1992).

[5] Lewis K., Trends Biochem. Sci. 19:119-123(1994).

15

852. (IBR)

In Between Ring fingers

The IBR (In Between Ring fingers) domain is found to occur between pairs of ring fingers (zf-C3HC4). The function of this domain is unknown. This domain has also been called the

20

C6HC domain and DRIL (for double RING finger linked) domain [2].

Number of members: 25

[1] Morett E, Bork P; Medline: 10366851 A novel transactivation domain in parkin."Trends Biochem Sci 1999;24:229-231.

25

[2] van der Reijden BA, Erpelinck-Verschueren CA, Lowenberg B, Jansen JH; Medline: 99349709 TRIADs: a new class of proteins with a novel cysteine-rich signature." Protein Sci 1999;8:1557-1561.

30

853. (IPPT)

IPP transferase

DDETF" 08668960

[1] Durand JM, Bjork GR, Kuwae A, Yoshikawa M, Sasakawa C; Medline: 97440126 The modified nucleoside 2-methylthio-N6-isopentenyladenosine in tRNA of *Shigella flexneri* is required for expression of virulence genes." J Bacteriol 1997;179:5777-5782.

[2] Boguta M, Hunter LA, Shen WC, Gillman EC, Martin NC, Hopper AK; Medline: 94187700 Subcellular locations of MOD5 proteins: mapping of sequences sufficient for targeting to mitochondria and demonstration that mitochondrial and nuclear isoforms commingle in the cytosol." Mol Cell Biol 1994;14:2298-2306.

[3] Gillman EC, Slusher LB, Martin NC, Hopper AK; Medline: 91203856 MOD5 translation initiation sites determine N6-isopentenyladenosine modification of mitochondrial and cytoplasmic tRNA." Mol Cell Biol 1991;11:2382-2390.

854. (KE2)

KE2 family protein

The function of members of this family is unknown, although they have been suggested to contain a DNA binding leucine zipper motif [2].

Number of members: 9

[1] Ha H, Abe K, Artzt K; Medline: 92084131 Primary structure of the embryo-expressed gene KE2 from the mouse H-2K region." Gene 1991;107:345-346.

[2] Shang HS, Wong SM, Tan HM, Wu M; Medline: 95129859 YKE2, a yeast nuclear gene encoding a protein showing homology to mouse KE2 and containing a putative leucine-zipper motif." Gene 1994;151:197-201.

855. (Lipoprotein_6)

Prokaryotic membrane lipoprotein lipid attachment site

Cross-reference(s) PS00013; PROKAR_LIPOPROTEIN

In prokaryotes, membrane lipoproteins are synthesized with a precursor signal peptide, which is cleaved by a specific lipoprotein signal peptidase (signal peptidase II). The peptidase recognizes a conserved sequence and cuts upstream of a cysteine residue to which

5

- 10

15

20

25

30

- *Vibrio harveyi* chitobiase (gene chb).
- *Yersinia* virulence plasmid protein yscJ.
- Halocyanin from *Natrobacterium pharaonis* [4], a membrane associated copper-binding protein. This is the first archaeobacterial protein known to be modified in such a fashion).

5

From the precursor sequences of all these proteins, a consensus pattern and a set of rules to identify this type of post-translational modification were derived.

Consensus pattern: {DERK}(6)-[LIVMFWSTAG](2)-[LIVMFYSTAGCQ]-[AGS]-C [C is
10 the lipid attachment site] Additional rules: 1)

The cysteine must be between positions 15 and 35 of the sequence in consideration. 2) There
must be at least one Lys or one Arg in the first seven positions of the sequence. Sequences
known to belong to this class detected by the pattern ALL. Other sequence(s) detected in
15 SWISS-PROT some 100 prokaryotic proteins. Some of them are not membrane lipoproteins,
but at least half of them could be.

References

- [1] Hayashi S., Wu H.C., J. Bioenerg. Biomembr. 22:451-471(1990).
- 20 [2] Klein P., Somorjai R.L., Lau P.C.K., Protein Eng. 2:15-20(1988).
- [3] von Heijne G., Protein Eng. 2:531-534(1989).
- [4] Mattar S., Scharf B., Kent S.B.H., Rodewald K., Oesterhelt D., Engelhard M. J. Biol.
Chem. 269:14939-14945(1994).

25

856. (Lipoprotein_7)

Adhesin lipoprotein

30

This family consists of the p50 and variable adherence-associated antigen (Vaa) adhesins
from *Mycoplasma hominis*. *M. hominis* is a mycoplasma associated with human urogenital
diseases, pneumonia, and septic arthritis [1]. An adhesin is a cell surface molecule that
mediates adhesion to other cells or to the surrounding surface or substrate. The Vaa antigen is
a 50-kDa surface lipoprotein that has four tandem repetitive DNA sequences encoding a
periodic peptide structure, and is highly immunogenic in the human host [1]. p50 is also a 50-

kDa lipoprotein, having three repeats A,B and C, that may be a tetramer of 191-kDa in its native environment [2].

Number of members: 18

5

[1] Zhang Q, Wise KS; Medline: 96294788 Molecular basis of size and antigenic variation of a *Mycoplasma hominis* adhesin encoded by divergent vaa genes. Infect Immun 1996;64:2737-2744.

10

[2] Henrich B, Kitzerow A, Feldmann RC, Schaal H, Hadding U; Medline: 97047675 Repetitive elements of the *Mycoplasma hominis* adhesin p50 can be differentiated by monoclonal antibodies." Infect Immun 1996;64:4027-4034.

857. (MaoC_like)

15

MaoC like domain

20

The MaoC protein is found to share similarity with a wide variety of enzymes; estradiol 17 beta-dehydrogenase 4, peroxisomal hydratase-dehydrogenase-epimerase, fatty acid synthase beta subunit. All these enzymes contain other domains. This domain is also present in the NodN nodulation protein N. No specific function has been assigned to this region of any of these proteins. The maoC gene is part of an operon with maoA which is involved in the synthesis of monoamine oxidase [1].

Number of members: 46

25

[1] Sugino H, Sasaki M, Azakami H, Yamashita M, Murooka Y Medline: 96235221 A monoamine-regulated *Klebsiella aerogenes* operon containing the monoamine oxidase structural gene (maoA) and the maoC gene." J Bacteriol 1992;174:2485-2492.

30

858. (MSP)

Manganese-stabilizing protein / photosystem II polypeptide

09689980-101300

This family consists of the 33 KDa photosystem II polypeptide from the oxygen evolving complex (OEC) of plants and cyanobacteria. The protein is also known as the manganese-stabilizing protein as it is associated with the manganese complex of the OEC and may provide the ligands for the complex [1].

Number of members: 17

[1] Philbrick JB, Zilinskas BA; Medline: 88334494 "Cloning, nucleotide sequence and mutational analysis of the gene encoding the Photosystem II manganese-stabilizing polypeptide of *Synechocystis* 6803." *Mol Gen Genet* 1988;212:418-425.

859. (NAC)

[1] Makarova KS, Aravind L, Galperin MY, Grishin NV, Tatusov RL, Wolf YI, Koonin EV; Medline: 99342100 Comparative genomics of the Archaea (Euryarchaeota): evolution of conserved protein families, the stable core, and the variable shell." *Genome Res* 1999;9:608-628.

Number of members: 27

860. (Nop)

Putative snoRNA binding domain

This family consists of various Pre RNA processing ribonucleoproteins. The function of the aligned region is unknown however it may be a common RNA or snoRNA or Nop1p binding domain. Nop5p (Nop58p) Swiss:Q12499 from yeast is the protein component of a ribonucleoprotein protein required for pre-18s rRNA processing and is suggested to function with Nop1p in a snoRNA complex [1]. Nop56p Swiss:O00567 and Nop5p interact with Nop1p and are required for ribosome biogenesis [2]. Prp31p Swiss:p49704 is required for pre-mRNA splicing in *S. cerevisiae* [3].

Number of members: 23

[1] Wu P, Brockenbrough JS, Metcalfe AC, Chen S, Aris JP; Medline: 98298165 Nop5p is a small nucleolar ribonucleoprotein component required for pre- 18 S rRNA processing in yeast." J Biol Chem 1998;273:16453-16463.

5 [2] Gautier T, Berges T, Tollervey D, Hurt E; Medline: 8038777 Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis." Mol Cell Biol 1997;17:7088-7098.

[3] Weidenhammer EM, Singh M, Ruiz-Noriega M, Woolford JL Jr; Medline: 96184869 The PRP31 gene encodes a novel protein required for pre-mRNA splicing in *Saccharomyces cerevisiae*." Nucleic Acids Res 1996;24:1164-1170.

861. (Nramp)

Natural resistance-associated macrophage protein

15 The natural resistance-associated macrophage protein (NRAMP) family consists of Nramp1, Nramp2, and yeast proteins Smf1 and Smf2. The NRAMP family is a novel family of functional related proteins defined by a conserved hydrophobic core of ten transmembrane domains [5]. This family of membrane proteins are divalent cation transporters. Nramp1 is an
20 integral membrane protein expressed exclusively in cells of the immune system and is recruited to the membrane of a phagosome upon phagocytosis [1]. By controlling divalent cation concentrations Nramp1 may regulate the interphagosomal replication of bacteria [1]. Mutations in Nramp1 may genetically predispose an individual to susceptibility to diseases including leprosy and tuberculosis conversely this might however provide protection from
25 rheumatoid arthritis [1]. Nramp2 is a multiple divalent cation transporter for Fe²⁺, Mn²⁺ and Zn²⁺ amongst others it is expressed at high levels in the intestine; and is major transferrin-independent iron uptake system in mammals [1]. The yeast proteins Smf1 and Smf2 may also transport divalent cations [3].

30 Number of members: 36

[1] Govoni G, Gros P; Medline: 98383996 Macrophage NRAMP1 and its role in resistance to microbial infections." Inflamm Res 1998;47:277-284.

[2] Agranoff DD, Krishna S Medline: 98294035 Metal ion homeostasis and intracellular parasitism." Mol Microbiol 1998;28:403-412.

[3] Pinner E, Gruenheid S, Raymond M, Gros P; Medline: 98030569 Functional complementation of the yeast divalent cation transporter family SMF by NRAMP2, a member of the mammalian natural resistance- associated macrophage protein family." J Biol Chem 1997;272:28933-28938.

[4] Cellier M, Belouchi A, Gros P; Medline: 96402487 Resistance to intracellular infections: comparative genomic analysis of Nramp." Trends Genet 1996;12:201-204.

[5] Cellier M, Prive G, Belouchi A, Kwan T, Rodrigues V, Chia W, Gros P; Medline: 96036029 Nramp defines a family of membrane proteins." Proc Natl Acad Sci U S A 1995;92:10089-10093.

862. (NTP_transf_2)

Nucleotidyltransferase domain

Members of this family belong to a large family of nucleotidyltransferases [1].

Number of members: 83

[1] Holm L, Sander C; Medline: 96005605 DNA polymerase beta belongs to an ancient nucleotidyltransferase superfamily." Trends Biochem Sci 1995;20:345-347.

863. (Paramyxo_P)

Paramyxovirus P phosphoprotein

This family consists of paramyxovirus P phosphoprotein from sendai virus and human and bovine parainfluenza viruses. The P protein is an essential part of the viral RNA polymerase complex formed from the P and L proteins [1]. The exact role of the P protein in this complex is unknown but it is involved in multiple protein-protein interactions and binding the polymerase complex to the nucleocapsid or ribonucleoprotein template [1]. It also appears to be important for the proper folding of the L protein [1]. The paramyxoviruses have a negative sense ssRNA genome [1].

Number of members: 15

[1] Bowman MC, Smallwood S, Moyer SA; Medline: 99329169 Dissection of Individual
5 Functions of the Sendai Virus Phosphoprotein in Transcription." J Virol 1999;73:6474-6483.
[2] Matsuoka Y, Curran J, Pelet T, Kolakofsky D, Ray R, Compans RW; Medline: 91237868
The P gene of human parainfluenza virus type 1 encodes P and C proteins but not a
cysteine-rich V protein." J Virol 1991;65:3406-3410.

10
864. (Patatin)

This family consists of various patatin glycoproteins from plants. The patatin protein
accounts for up to 40% of the total soluble protein in potato tubers [2]. Patatin is a storage
15 protein but it also has the enzymatic activity of lipid acyl hydrolase, catalysing the cleavage
of fatty acids from membrane lipids [2].

Number of members: 21

[1] Banfalvi Z, Kostyal Z, Barta E; Medline: 95107249 Solanum brevidens possesses a non-
20 sucrose-inducible patatin gene." Mol Gen Genet 1994;245:517-522.
[2] Mignery GA, Pikaard CS, Park WD; Medline: 88226014 Molecular characterization of
the patatin multigene family of potato." Gene 1988;62:27-44.

25
865. (Pentapeptide_2)

Pentapeptide repeats (8 copies)

These repeats are found in many mycobacterial proteins. These repeats are most common in
30 the PPE family of proteins, where they are found in the MPTR subfamily of PPE proteins.
The function of these repeats is unknown. The repeat can be approximately described as
XNXGX, where X can be any amino acid. These repeats are similar to Pentapeptide [1],
however it is not clear if these two families are structurally related.

Number of members: 362

5

10

Peptidase C13 family

15

20

25

Proline dehydrogenase

30

[1] Ling M, Allen SW, Wood JM; Medline: 95055736 Sequence analysis identifies the proline dehydrogenase and delta 1- pyrroline-5-carboxylate dehydrogenase domains of the multifunctional Escherichia coli PutA protein." J Mol Biol 1994;243:950-956.

5

868. (PsbP)

This family consists of the 23 kDa subunit of oxygen evolving system of photosystem II or PsbP from various plants (where it is encoded by the nuclear genome) and Cyanobacteria.

10

The 23 KDa PsbP protein is required for PSII to be fully operational in vivo, it increases the affinity of the water oxidation site for Cl- and provides the conditions required for high affinity binding of Ca²⁺ [2].

Number of members: 25

15

[1] Rova EM, Mc Ewen B, Fredriksson PO, Styring S; Medline: 97067138 Photoactivation and photoinhibition are competing in a mutant of Chlamydomonas reinhardtii lacking the 23-kDa extrinsic subunit of photosystem II." J Biol Chem 1996;271:28918-28924.

20

[2] Kochhar A, Khurana JP, Tyagi AK; Medline: 97191538 Nucleotide sequence of the psbP gene encoding precursor of 23-kDa polypeptide of oxygen-evolving complex in Arabidopsis thaliana and its expression in the wild-type and a constitutively photomorphogenic mutant." DNA Res 1996;3:277-285.

25

869. (PUA)

The PUA domain named after PseudoUridine synthase and Archaeosine transglycosylase, was detected in archaeal and eukaryotic pseudouridine synthases, archaeal archaeosine synthases, a family of predicted ATPases that may be involved in RNA modification, a

30

family of predicted archaeal and bacterial rRNA methylases. Additionally, the PUA domain was detected in a family of eukaryotic proteins that also contain a domain homologous to the translation initiation factor eIF1/SUI1; these proteins may comprise a novel type of translation factors. Unexpectedly, the PUA domain was detected also in bacterial and yeast glutamate kinases; this is compatible with the demonstrated role of these enzymes in the

regulation of the expression of other genes [1]. It is predicted that the PUA domain is an RNA binding domain.

Number of members: 48

[1] Aravind L, Koonin EV; Medline: 99193178 Novel predicted RNA-binding domains associated with the translation machinery." J Mol Evol 1999;48:291-302.

870. (RF1)
eRF1-like proteins

Members of this family are peptide chain release factors. The eukaryotic Release Factor 1 proteins (eRF1s) are involved in termination of translation. The eRF1 protein is functional for all stop codons and appears to abolish read-through of these codons. This family also includes other proteins for which the precise molecular function is unknown. Many of them are from Archaeobacteria. These proteins may also be involved in translation termination but this awaits experimental verification. Number of members: 25

[1] Frolova L, Le Goff X, Rasmussen HH, Cheperegin S, Drugeon G, Kress M, Arman I, Haenni AL, Celis JE, Philippe M, et al; Medline: 95082951 A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor” [see comments] Nature 1994;372:701-703.

[2] Drugeon G, Jean-Jean O, Frolova L, Le Goff X, Philippe M, Kisselev L, Haenni AL; Medline: 97315314 Eukaryotic release factor 1 (eRF1) abolishes readthrough and competes with suppressor tRNAs at all three termination codons in messenger RNA.” Nucleic Acids Res 1997;25:2254-2258.

871. (Ribosomal_L14e)Ribosomal protein L14
This family includes the eukaryotic ribosomal protein L14.
Number of members: 15

872. (Ribosomal_S27)

Ribosomal protein S27a

This family of ribosomal proteins consists mainly of the 40S ribosomal protein S27a which is synthesized as a C-terminal extension of ubiquitin (CEP). The S27a domain compromises the C-terminal half of the protein. The synthesis of ribosomal proteins as extensions of ubiquitin promotes their incorporation into nascent ribosomes by a transient metabolic stabilization and is required for efficient ribosome biogenesis [3]. The ribosomal extension protein S27a contains a basic region that is proposed to form a zinc finger; its fusion gene is proposed as a mechanism to maintain a fixed ratio between ubiquitin necessary for degrading proteins and ribosomes a source of proteins [2].

Number of members: 36

873. (Spermine_synth)

Spermine/spermidine synthase

Spermine and spermidine are polyamines. This family includes spermidine synthase that catalyses the fifth (last) step in the biosynthesis of spermidine from arginine, and spermine synthase.

Number of members: 39

[1] Mezquita J, Pau M, Mezquita C; Medline: 97449308 Characterization and expression of two chicken cDNAs encoding ubiquitin fused to ribosomal proteins of 52 and 80 amino acids." Gene 1997;195:313-319.

[2] Redman KL, Rechsteiner M; Medline: 89181932 Identification of the long ubiquitin extension as ribosomal protein S27a." Nature 1989;338:438-440.

[3] Finley D, Bartel B, Varshavsky A; Medline: 89181925 The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis." Nature 1989;338:394-401.

874. (Surp)

Surp module

[1] Denhez F, Lafyatis R; Medline: 94266805 Conservation of regulated alternative splicing and identification of functional domains in vertebrate homologs to the Drosophila splicing regulator, suppressor-of-white-apricot." J Biol Chem 1994;269:16170-16179.

This domain is also known as the SWAP domain. SWAP stands for Suppressor-of-White-APricot. It has been suggested that these domains may be RNA binding [1].

Number of members: 32

875. (TFIIE)

TFIIE alpha subunit

The general transcription factor TFIIE has an essential role in eukaryotic transcription initiation together with RNA polymerase II and other general factors. Human TFIIE consists of two subunits TFIIE-alpha Swiss:P29083 and TFIIE-beta Swiss:P29084 and joins the preinitiation complex after RNA polymerase II and TFIIF [1]. This family consists of the conserved amino terminal region of eukaryotic TFIIE-alpha [2] and proteins from archaeobacteria that are presumed to be TFIIE-alpha subunits also Swiss:O29501 [3].

Number of members: 12

[1] Ohkuma Y, Sumimoto H, Hoffmann A, Shimasaki S, Horikoshi M, Roeder RG; Medline: 92065982 Structural motifs and potential sigma homologies in the large subunit of human general transcription factor TFIIE." Nature 1991;354:398-401.

[2] Ohkuma Y, Hashimoto S, Roeder RG, Horikoshi M; Medline: 93087200 Identification of two large subdomains in TFIIE-alpha on the basis of homology between Xenopus and human sequences. Nucleic Acids Res 1992;20:5838-5838.

[3] Klenk HP, Clayton RA, Tomb JF, White O, Nelson KE, Ketchum KA, Dodson RJ, Gwinn M, Hickey EK, Peterson JD, Richardson DL, Kerlavage AR, Graham DE, Kyrpides NC, Fleischmann RD, Quackenbush J, Lee NH, Sutton GG, Gill S, Kirkness EF, Dougherty BA,

876. (Transglut core)

Cross-reference(s) PS00547; TRANSGLUTAMINASES

Transglutaminases (EC 2.3.2.13) (TGase) [1,2] are calcium-dependent enzymes that catalyze the cross-linking of proteins by promoting the formation of isopeptide bonds between the gamma-carboxyl group of a glutamine in one polypeptide chain and the epsilon-amino group of a lysine in a second polypeptide chain. TGases also catalyze the conjugation of polyamines to proteins. The best known transglutaminase is blood coagulation factor XIII, a plasma tetrameric protein composed of two catalytic A subunits and two non-catalytic B subunits. Factor XIII is responsible for cross-linking fibrin chains, thus stabilizing the fibrin clot. Other forms of transglutaminases are widely distributed in various organs, tissues and body fluids. Sequence data is available for the following forms of TGase:

- 20 - Transglutaminase K (Tgase K), a membrane-bound enzyme found in mammalian epidermis
and important for the formation of the cornified cell envelope (gene TGM1).
- Tissue transglutaminase (TGase C), a monomeric ubiquitous enzyme located in the
cytoplasm (gene TGM2).
- Transglutaminase 3, responsible for the later stages of cell envelope formation in the
epidermis and the hair follicle (gene TGM3).
25 - Transglutaminase 4 (gene TGM4).

A conserved cysteine is known to be involved in the catalytic mechanism of TGases. The erythrocyte membrane band 4.2 protein, which probably plays an important role in regulating the shape of erythrocytes and their mechanical properties, is evolutionary related to TGases. However the active site cysteine is substituted by an alanine and the 4.2 protein does not show TGase activity.

Consensus pattern:[GT]-Q-[CA]-W-V-x-[SA]-[GA]-[IVT]-x(2)-T-x-[LMSC]-R-[CSA]-[LV]-G [The first C is the active site residue] Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROT NONE.

- 5 [1] Ichinose A., Bottenus R.E., Davie E.W. J. Biol. Chem. 265:13411-13414(1990).
[2] Greenberg C.S., Birckbichler P.J., Rice R.H. FASEB J. 5:3071-3077(1991).

877. (TruB_N)

10 TruB family pseudouridylate synthase (N terminal domain)

Members of this family are involved in modifying bases in RNA molecules. They carry out the conversion of uracil bases to pseudouridine. This family includes TruB, a pseudouridylate synthase that specifically converts uracil 55 to pseudouridine in most tRNAs. This family
15 also includes Cbf5p that modifies rRNA [2].

Number of members: 33

[1] Nurse K, Wrzesinski J, Bakin A, Lane BG, Ofengand J; Medline: 96079944 Purification, cloning, and properties of the tRNA psi 55 synthase from Escherichia coli." RNA
20 1995;1:102-112.

[2] Lafontaine DLJ, Bousquet-Antonelli C, Henry Y, Caizergues-Ferrer M, Tollervey D; Medline: 98139521 The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase." Genes Dev 1998;12:527-537.

878. (UDPGP)

UTP--glucose-1-phosphate uridylyltransferase

30 This family consists of UTP--glucose-1-phosphate uridylyltransferases, EC:2.7.7.9. Also known as UDP-glucose pyrophosphorylase (UDPGP) and Glucose-1-phosphate uridylyltransferase. UTP--glucose-1-phosphate uridylyltransferase catalyses the interconversion of MgUTP + glucose-1-phosphate and UDP-glucose + MgPPi [1]. UDP-glucose is an important intermediate in mammalian carbohydrate interconversion involved in

5

Number of members: 18

10

[2] Ragheb JA, Dottin RP; Medline: 87231075 Structure and sequence of a UDP glucose pyrophosphorylase gene of Dictyostelium discoideum.” Nucleic Acids Res 1987;15:3891-3906.

15

20

25

- *Bacillus subtilis* hypothetical protein yqeI.
- *Escherichia coli* hypothetical protein yhbY and HI1333, the corresponding *Haemophilus influenzae* protein.
- *Methanococcus jannaschii* hypothetical protein MJ0652.

30

Consensus pattern: L-[ST]-x(3)-K-x(3)-[KR]-[SGA]-x-[GA]-H-x-L-x-P-[LIV]-x(2)-[LIV]-[GA]-x(2)-G Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROT/NONE.

5

880. (zf-A20)

A20-like zinc finger

A20- (an inhibitor of cell death)-like zinc fingers. The zinc finger mediates self-association in A20. These fingers also mediate IL-1-induced NF-kappa B activation.

10

Number of members: 22

[1] Heyninck K, Beyaert R; Medline: 99126071 The cytokine-inducible zinc finger protein A20 inhibits IL-1-induced NF- kappaB activation at the level of TRAF6. FEBS Lett 1999;442:147-150.

15

[2] De Valck D, Heyninck K, Van Crielinge W, Contreras R, Beyaert R, Fiers W; Medline: 96390831 A20, an inhibitor of cell death, self-associates by its zinc finger domain." FEBS Lett 1996;384:61-64.

20

[3] Song HY, Rothe M, Goeddel DV; Medline: 96270609 The tumor necrosis factor-inducible zinc finger protein A20 interacts with TRAF1/TRAF2 and inhibits NF-kappaB activation. Proc Natl Acad Sci U S A 1996;93:6721-6725.

[4] Opipari AW Jr, Boguski MS, Dixit VM; Medline: 90368626 The A20 cDNA induced by tumor necrosis factor alpha encodes a novel type of zinc finger protein." J Biol Chem 1990;265:14705-14708.

25

881. (zf-PARP)

Poly(ADP-ribose) polymerase zinc finger domain

30

Cross-reference(s) PS00347; PARP_ZN_FINGER_1 PS50064; PARP_ZN_FINGER_2

Poly(ADP-ribose) polymerase (EC 2.4.2.30) (PARP) [1,2] is a eukaryotic enzyme that catalyzes the covalent attachment of ADP-ribose units from NAD(+) to various nuclear

0965990-101300

5

15

20

- 25

30

Enzyme that catalyses the phosphorylation of adenylylsulfate to 3'-phosphoadenylylsulfate. This domain contains an ATP binding P-loop motif. Number of members: 34

[1] MacRae IJ, Rose AB, Segel IH; Medline: 99003196 Adenosine 5'-phosphosulfate kinase from *Penicillium chrysogenum*. site- directed mutagenesis at putative phosphoryl-accepting and ATP P-loop residues. J Biol Chem 1998;273:28583-28589.

5 883. DNA polymerase family B signature DNA_POLYMERASE_B (DNA_pol_B)

Replicative DNA polymerases (EC 2.7.7.7) are the key enzymes catalyzing the accurate replication of DNA. They require either a small RNA molecule or a protein as a primer for the de novo synthesis of a DNA chain. On the basis of sequence similarity, a number of DNA polymerases have been grouped [1 to 7] under the designation of DNA polymerase family B. These are:

- Higher eukaryotes polymerases alpha.
- Higher eukaryotes polymerases delta.
- Yeast polymerase I/alpha (gene POL1), polymerase II/epsilon (gene POL2), polymerase III/delta (gene POL3) and polymerase REV3.
- *Escherichia coli* polymerase II (gene *dinA* or *polB*).
- Archaeobacterial polymerases.
- Polymerases of viruses from the herpesviridae family.
- Polymerases from Adenoviruses.
- Polymerases from Baculoviruses.
- Polymerases from Chlorella viruses.
- Polymerases from Poxviruses.
- Bacteriophage T4 polymerase.
- Podoviridae bacteriophages Phi-29, M2 and PZA polymerase.
- Tectiviridae bacteriophage PRD1 polymerase.
- Polymerases encoded on mitochondrial linear DNA plasmids in various fungi and plants (*Kluyveromyces lactis* pGKL1 and pGKL2, *Agaricus bitorquis* pEM, *Ascobolus immersus* pAI2, *Claviceps purpurea* pCLK1, *Neurospora Kalilo* and *Maranhar*, maize S-1, etc).

30 Six regions of similarity (numbered from I to VI) are found in all or a subset of the above polymerases. The most conserved region (I) includes a conserved tetrapeptide with two aspartate residues. Its function is not yet known. However, it has been suggested [3] that it may be involved in binding a magnesium ion. This conserved region was selected as a signature for this family of DNA polymerases.

Consensus pattern [YA]-[GLIVMSTAC]-D-T-D-[SG]-[LIVMFTC]-x-[LIVMSTAC]
Sequences known to belong to this class detected by the patternALL, except for yeast
polymerase II/epsilon, Agaricus bitorquis pEM and Sulfolobus solfataricus polymerase II.

- [1] Jung G., Leavitt M.C., Hsieh J.-C., Ito J. Proc. Natl. Acad. Sci. U.S.A. 84:8287-8291(1987).
[2] Bernad A., Zaballos A., Salas M., Blanco L. EMBO J. 6:4219-4225(1987).
[3] Argos P. Nucleic Acids Res. 16:9909-9916(1988).
[4] Wang T.S.-F., Wong S.W., Korn D. FASEB J. 3:14-21(1989).
[5] Delarue M., Poch O., Todro N., Moras D., Argos P. Protein Eng. 3:461-467(1990).
[6] Ito J., Braithwaite D.K. Nucleic Acids Res. 19:4045-4057(1991).
[7] Braithwaite D.K., Ito J. Nucleic Acids Res. 21:787-802(1993).

884. DNA polymerase family X signature - DNA_POLYMERASE_X (DNA_polymeraseX)

DNA polymerases (EC 2.7.7.7) can be classified, on the basis of sequence similarity [1], into
at least four different groups: A, B, C and X. DNA polymerases that belong to family X are
listed below [2]:

- Vertebrate polymerase beta, involved in DNA repair.
- Yeast polymerase IV (POL4) [3], an enzyme with similar characteristics to that of the
mammalian polymerase beta.
- Terminal deoxynucleotidyltransferase (TdT) (EC 2.7.7.31). TdT catalyzes the elongation of
polydeoxynucleotide chains by terminal addition. One of the functions of this enzyme is the
addition of nucleotides at the junction of rearranged Ig heavy chain and T cell receptor gene
segments during the maturation of B and T cells.
- African Swine Fever virus protein O174L [4].
- Fission yeast hypothetical protein SpAC2F7.06c.

These enzymes are small (about 40 Kd) compared with other polymerases and their reaction
mechanism operates via a distributive mode, i.e. they dissociate from the template-primer
after addition of each nucleotide.

As a signature pattern for this family of DNA polymerases, a highly conserved region that contains a conserved arginine and two conserved aspartic acid residues were selected. The latter together with the arginine have been shown [5] to be involved in primer binding in polymerase beta.

Consensus pattern G-[SG]-[LFY]-x-R-[GE]-x(3)-[SGCL]-x-D-[LIVM]-D-[LIVMFY](3)-x(2)-[SAP] Sequences known to belong to this class detected by the patternALL.

[1] Ito J., Braithwaite D.K. Nucleic Acids Res. 19:4045-4057(1991).

[2] Matsukage A., Nishikawa K., Ooi T., Seto Y., Yamaguchi M. J. Biol. Chem. 262:8960-8962(1987).

[3] Prasad R., Widen S.G., Singhal R.K., Watkins J., Prakash L., Wilson S.H. Nucleic Acids Res. 21:5301-5307(1993).

[4] Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C., Rodriguez J.F., Vinuela E. Virology 208:249-278(1995).

[5] Date T., Yamamoto S., Tanihara K., Nishimoto Y., Matsukage A. Biochemistry 30:5286-5292(1991).

885. DUF14 - Domain of unknown function

This domain is found in glutamate synthase, tungsten formylmethanofuran dehydrogenase subunit c (FwdC) and molybdenum formylmethanofuran dehydrogenase subunit c (FmdC). It has no known function. Number of members: 52

[1] Hochheimer A, Hedderich R, Thauer RK; Medline: 99035764. The formylmethanofuran dehydrogenase isoenzymes in *Methanobacterium wolfei* and *Methanobacterium thermoautotrophicum*: induction of the molybdenum isoenzyme by molybdate and constitutive synthesis of the tungsten isoenzyme." Arch Microbiol 1998;170:389-393.

886. DUF18-Domain of unknown function

This domain of unknown function is found in several *C. elegans* proteins. The domain is 120 amino acids long and rich in cysteine residues. There are 16 conserved cysteine positions in the domain. Number of members: 34

887. DUF27-Domain of unknown function

5

10

15

20

25

30

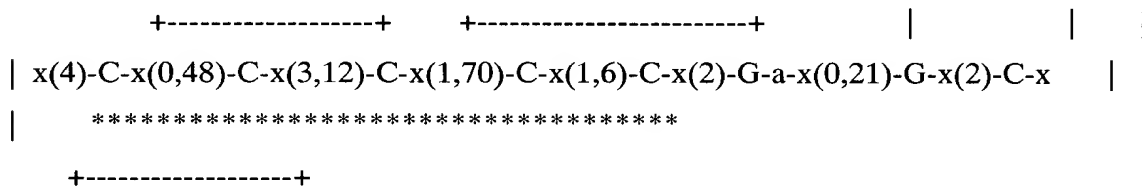
- Calcium-dependent serine proteinase (CASP) which degrades the extracellular matrix proteins type I and IV collagen and fibronectin (1 copy).

- Cartilage matrix protein CMP (1 copy).
- Cartilage oligomeric matrix protein COMP (4 copies).
- Cell surface antigen 114/A10 (3 copies).
- Cell surface glycoprotein complex transmembrane subunit ASGP-2 from rat (2 copies).
- 5 - Coagulation associated proteins C, Z (2 copies) and S (4 copies).
- Coagulation factors VII, IX, X and XII (2 copies).
- Complement C1r components (1 copy).
- Complement C1s components (1 copy).
- Complement-activating component of Ra-reactive factor (RARF) (1 copy).
- 10 - Complement components C6, C7, C8 alpha and beta chains, and C9 (1 copy).
- Crumbs, an epithelial development protein from Drosophila (29 copies).
- Epidermal growth factor precursor (7-9 copies).
- Exogastrula-inducing peptides A, C, D and X from sea urchin (1 copy).
- Fat protein, a Drosophila cadherin-related tumor suppressor (5 copies).
- 15 - Fetal antigen 1, a probable neuroendocrine differentiation protein, which is derived from the delta-like protein (DLK) (6 copies).
- Fibrillin 1 (47 copies) and fibrillin 2 (14 copies).
- Fibropellins IA (21 copies), IB (13 copies), IC (8 copies), II (4 copies) and III (8 copies) from the apical lamina - a component of the extracellular matrix - of sea urchin.
- 20 - Fibulin-1 and -2, two extracellular matrix proteins (9-11 copies).
- Giant-lens protein (protein Argos), which regulates cell determination and axon guidance in the Drosophila eye (1 copy).
- Growth factor-related proteins from various poxviruses (1 copy).
- Gurken protein, a Drosophila developmental protein (1 copy).
- 25 - Heparin-binding EGF-like growth factor (HB-EGF), transforming growth factor alpha (TGF-alpha), growth factors Lin-3 and Spitz (1 copy); the precursors are membrane proteins, the mature form is located extracellular.
- Hepatocyte growth factor (HGF) activator (EC 3.4.21.-) (2 copies).
- LDL and VLDL receptors, which bind and transport low-density lipoproteins and very low-
- 30 density lipoproteins (3 copies).
- LDL receptor-related protein (LRP), which may act as a receptor for endocytosis of extracellular ligands (22 copies).
- Leucocyte antigen CD97 (3 copies), cell surface glycoprotein EMR1 (6 copies) and cell surface glycoprotein F4/80 (7 copies).

- Thrombomodulin (fetomodulin), which together with thrombin activates protein C (6 copies).

- Thrombospondin 1, 2 (3 copies), 3 and 4 (4 copies), adhesive glycoproteins that mediate cell-to-cell and cell-to-matrix interactions.
- Thyroid peroxidase 1 and 2 (EC 1.11.1.8) from human (1 copy).
- Transforming growth factor beta-1 binding protein (TGF-B1-BP) (16 or 18 copies).
- Tyrosine-protein kinase receptors Tek and Tie (EC 2.7.1.112) (3 copies).
- Urokinase-type plasminogen activator (EC 3.4.21.73) (UPA) and tissue plasminogen activator (EC 3.4.21.68) (TPA) (1 copy).
- Uromodulin (Tamm-horsfall urinary glycoprotein) (THP) (3 copies).
- Vitamin K-dependent anticoagulants protein C (2 copies) and protein S (4 copies) and the similar protein Z, a single-chain plasma glycoprotein of unknown function (2 copies).
- 63 Kd sperm flagellar membrane protein from sea urchin (3 copies).
- 93 Kd protein (gene nel) from chicken (5 copies).
- Hypothetical 337.6 Kd protein T20G5.3 from *Caenorhabditis elegans* (44 copies).

The functional significance of EGF domains in what appear to be unrelated proteins is not yet clear. However, a common feature is that these repeats are found in the extracellular domain of membrane-bound proteins or in proteins known to be secreted (exception: prostaglandin G/H synthase). The EGF domain includes six cysteine residues which have been shown (in EGF) to be involved in disulfide bonds. The main structure is a two-stranded beta-sheet followed by a loop to a C-terminal short two-stranded sheet. Subdomains between the conserved cysteines strongly vary in length as shown in the following schematic representation of the EGF-like domain:



'C': conserved cysteine involved in a disulfide bond.

'G': often conserved glycine

'a': often conserved aromatic amino acid

'*': position of both patterns.

'x': any residue

The region between the 5th and 6th cysteine contains two conserved glycines of which at least one is present in most EGF-like domains. Two patterns were created for this domain, each including one of these C-terminal conserved glycine residues.

5 Consensus pattern: C-x-C-x(5)-G-x(2)-C [The 3 C's are involved in disulfide bonds]
Sequences known to belong to this class detected by the pattern A majority, but not those that
have very long or very short regions between the last 3 conserved cysteines of their EGF-like
domain(s). Other sequence(s) detected in SWISS-PROT87 proteins, of which 27 can be
considered as possible candidates.

10

15

20

25

30

890. Ham1 family (Ham1p_like)

[1] Noskov VN, Staak K, Shcherbakova PV, Kozmin SG, Negishi K, Ono BC, Hayatsu H, Pavlov YI; Medline: 96381244 HAM1, the gene controlling 6-N-hydroxylaminopurine sensitivity and mutagenesis in the yeast *Saccharomyces cerevisiae*.” *Yeast* 1996;12:17-29.

Anion exchange is a cellular transport function which contributes to the regulation of cell pH and volume. Anion exchangers are a family of functionally related proteins that contributes to these properties by maintaining the intracellular level of the two principal anions: chloride and HCO_3^- . The best characterized anion exchanger is the band 3 protein [1], which is an erythrocyte anion exchange membrane glycoprotein. Band 3 is a protein of about 900 amino acids which consists of a cytoplasmic N-terminal domain of about 400 residues and an hydrophobic C-terminal section of about 500 residues that contains at least ten transmembrane regions. The cytoplasmic domain provides binding sites for cytoskeletal proteins, while the integral membrane domain is responsible for anion transport. Band 3 protein is specific to erythroid cells, at least two other proteins [2] structurally and functionally related to band 3, are found in nonerythroid tissues:

- Structurally AE2 and AE3 are very similar to band 3, the main difference being an extension of some 300 residues of the N-terminal domain in AE2 and AE3.

Two signature patterns were developed for these proteins. The first pattern is based on a conserved stretch of sequence that contains four clustered positive charged residues and which is located at the C-terminal extremity of the cytoplasmic domain, just before the first transmembrane segment from the integral domain. The second pattern is based on the perfectly conserved sequence of the fifth transmembrane segment; this segment contains a lysine, which is the covalent binding site for the isothiocyanate group of DIDS, an inhibitor of anion exchange.

Consensus pattern F-G-G-[LIVM](2)-[KR]-D-[LIVM]-[RK]-R-R-Y Sequences known to belong to this class detected by the pattern ALL.

Consensus pattern [FI]-L-I-S-L-I-F-I-Y-E-T-F-x-K-L Sequences known to belong to this class detected by the pattern ALL.

- [1] Jay D., Cantley L. Annu. Rev. Biochem. 55:511-538(1986).
[2] Reithmeier R.A.F. Curr. Opin. Struct. Biol. 3:515-523(1993).

892. ATP phosphoribosyltransferase signature (HisG)

ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern a region located in the C-terminal part of this enzyme was selected.

Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-G-x-T-[LM]
Sequences known to belong to this class detected by the pattern ALL.

893. HNH endonuclease (HNH)

Number of members: 56

- [1] Shub DA, Goodrich-Blair H, Eddy SR; Medline: 95117127 Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns." Trends Biochem Sci 1994;19:402-404.

- [2] Dalgaard JZ, Klar AJ, Moser MJ, Holley WR, Chatterjee A, Mian IS; Medline: 98026854 Statistical modeling and analysis of the LAGLIDADG family of site- specific endonucleases

and identification of an intein that encodes a site-specific endonuclease of the HNH family.”
Nucleic Acids Res 1997;25:4626-4638.

[3] Gorbalenya AE; Medline: 95004046 Self-splicing group I and group II introns encode
homologous (putative) DNA endonucleases of a new family.” Protein Sci 1994;3:1117-1120.

5
894. NEUROHYPOPHYS_HORM (hormone5)

Oxytocin (or ocytocin) and vasopressin [1] are small (nine amino acid residues), structurally
and functionally related neurohypophysial peptide hormones. Oxytocin causes contraction of
the smooth muscle of the uterus and of the mammary gland while vasopressin has a direct
antidiuretic action on the kidney and also causes vasoconstriction of the peripheral vessels.
Like the majority of active peptides, both hormones are synthesized as larger protein
precursors that are enzymatically converted to their mature forms. Peptides belonging to this
family are also found in birds, fish, reptiles and amphibians (mesotocin, isotocin, valitocin,
glumitocin, aspartocin, vasotocin, seritocin, asvatocin, phasvatocin), in worms (annetocin),
octopi (cephalotocin), locust (locupressin or neuropeptide F1/F2) and in molluscs
(conopressins G and S) [2]. The pattern developed to detect this category of peptides spans
their entire sequence and includes four invariant amino acid residues.

Consensus pattern C-[LIFY](2)-x-N-[CS]-P-x-G [The two C's are linked by a disulfide
bond]. Sequences known to belong to this class detected by the pattern ALL.

[1] Acher R., Chauvet J. Biochimie 70:1197-1207(1988).

[2] Chauvet J., Michel G., Ouedraogo Y., Chou J., Chait B.T., Acher R. Int. J. Pept. Protein
Res. 45:482-487(1995).

895. 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase (HPPK)

All organisms require reduced folate cofactors for the synthesis of a variety of metabolites.
Most microorganisms must synthesize folate de novo because they lack the active transport
system of higher vertebrate cells which allows these organisms to use dietary folates.
Enzymes involved in folate biosynthesis are therefore targets for a variety of antimicrobial
agents such as trimethoprim or sulfonamides. 7,8-dihydro-6-hydroxymethylpterin-
pyrophosphokinase (EC 2.7.6.3) (HPPK) catalyzes the attachment of pyrophosphate to 6-
hydroxymethyl-7,8-dihydropterin to form 6-hydroxymethyl-7,8-dihydropteridine

pyrophosphate. This is the first step in a three-step pathway leading to 7,8-dihydrofolate. Bacterial HPPK (gene folK or sulD) [1] is a protein of 160 to 270 amino acids. In the lower eukaryote *Pneumocystis carinii*, HPPK is the central domain of a multifunctional folate synthesis enzyme (gene fas) [2]. As a signature for HPPK, a conserved region located in the central section of these enzymes was selected.

Consensus pattern [KRHD]-x-[GA]-[PSAE]-R-x(2)-D-[LIV]-D-[LIVM](2) Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROTNONE.

[1] Talarico T.L., Ray P.H., Dev I.K., Merrill B.M., Dallas W.S. J. Bacteriol. 174:5971-5977(1992).

[2] Volpes F., Dyer M., Scaife J.G., Darby G., Stammers D.K., Delves C.J. Gene 112:213-218(1992).

896. Metalloenzyme superfamily (Metalloenzyme)

This family includes phosphopentomutase Swiss:P07651 and 2,3-bisphosphoglycerate-independent phosphoglycerate mutase, Swiss:P37689. This family is also related to alk_phosphatase [1]. The alignment contains the most conserved residues that are probably involved in metal binding and catalysis. Number of members: 34

[1] Galperin MY, Bairoch A, Koonin EV; Medline: 99180418 A superfamily of metalloenzymes unifies phosphopentomutase and cofactor- independent phosphoglycerate mutase with alkaline phosphatases and sulfatases." Protein Sci 1998;7:1829-1835.

897. Penicillin amidase (Penicil_amidase)

Penicillin amidase or penicillin acylase EC:3.5.1.11 catalyses the hydrolysis of benzylpenicillin to phenylacetic acid and 6-aminopenicillanic acid (6-APA) a key intermediate in the the synthesis of penicillins [1]. Also in the family is cephalosporin acylase Swiss:P07662 and Swiss:P29958 aculeacin A acylase which are involved in the synthesis of related peptide antibiotics. Number of members: 13

[1] Verhaert RM, Riemens AM, van der Laan JM, van Duin J, Quax WJ; Medline: 97438505
Molecular cloning and analysis of the gene encoding the thermostable penicillin G acylase
from *Alcaligenes faecalis*. *Appl Environ Microbiol* 1997;63:3412-3418.

[2] Duggleby HJ, Tolley SP, Hill CP, Dodson EJ, Dodson G, Moody PC; Medline: 95115804
5 Penicillin acylase has a single-amino-acid catalytic centre." *Nature* 1995;373:264-268.

898. Phosphoribosyl-AMP cyclohydrolase (PRA-CH)

This enzyme catalyses the third step in the histidine biosynthetic pathway. It requires Zn ions
10 for activity. Number of members: 13

[1] D'Ordine RL, Klem TJ, Davisson VJ; Medline: 99129952 N1-(5'-
phosphoribosyl)adenosine-5'-monophosphate cyclohydrolase: purification and
characterization of a unique metalloenzyme. *Biochemistry* 1999;38:1537-1546.

899. Phosphoribosyl-ATP pyrophosphohydrolase (PRA-PH)

This enzyme catalyses the second step in the histidine biosynthetic pathway. Number of
members: 32

[1] Keesey JK Jr, Bigelis R, Fink GR; Medline: 79216449 The product of the *his4* gene
cluster in *Saccharomyces cerevisiae*. A trifunctional polypeptide." *J Biol Chem* 1979 Aug
10;254:7427-7433.

[2] Bruni CB, Carlomagno MS, Formisano S, Paoletta G; Medline: 86310274 Primary and
25 secondary structural homologies between the *HIS4* gene product of *Saccharomyces*
cerevisiae and the *hisIE* and *hisD* gene products of *Escherichia coli* and *Salmonella*
typhimurium." *Mol Gen Genet* 1986;203:389-396.

900. Prokaryotic membrane lipoprotein lipid attachment site (PstS)

In prokaryotes, membrane lipoproteins are synthesized with a precursor signal peptide, which
is cleaved by a specific lipoprotein signal peptidase (signal peptidase II). The peptidase
recognizes a conserved sequence and cuts upstream of a cysteine residue to which a

[illegible]

- Major outer membrane lipoprotein (murein-lipoproteins) (gene lpp).
- Escherichia coli lipoprotein-28 (gene nlpA).
- 5 - Escherichia coli lipoprotein-34 (gene nlpB).
- Escherichia coli lipoprotein nlpC.
- Escherichia coli lipoprotein nlpD.
- Escherichia coli osmotically inducible lipoprotein B (gene osmB).
- Escherichia coli osmotically inducible lipoprotein E (gene osmE).
- 10 - Escherichia coli peptidoglycan-associated lipoprotein (gene pal).
- Escherichia coli rare lipoproteins A and B (genes rplA and rplB).
- Escherichia coli copper homeostasis protein cutF (or nlpE).
- Escherichia coli plasmids traT proteins.
- Escherichia coli Col plasmids lysis proteins.
- 15 - A number of Bacillus beta-lactamases.
- Bacillus subtilis periplasmic oligopeptide-binding protein (gene oppA).
- Borrelia burgdorferi outer surface proteins A and B (genes ospA and ospB).
- Borrelia hermsii variable major protein 21 (gene vmp21) and 7 (gene vmp7).
- Chlamydia trachomatis outer membrane protein 3 (gene omp3).
- 20 - Fibrobacter succinogenes endoglucanase cel-3.
- Haemophilus influenzae proteins Pal and Pcp.
- Klebsiella pullulunase (gene pulA).
- Klebsiella pullulunase secretion protein pulS.
- Mycoplasma hyorhinitis protein p37.
- 25 - Mycoplasma hyorhinitis variant surface antigens A, B, and C (genes vlpABC).
- Neisseria outer membrane protein H.8.
- Pseudomonas aeruginosa lipopeptide (gene lppL).
- Pseudomonas solanacearum endoglucanase egl.
- Rhodopseudomonas viridis reaction center cytochrome subunit (gene cytC).
- 30 - Rickettsia 17 Kd antigen.
- Shigella flexneri invasion plasmid proteins mxiJ and mxiM.
- Streptococcus pneumoniae oligopeptide transport protein A (gene amiA).
- Treponema pallidum 34 Kd antigen.
- Treponema pallidum membrane protein A (gene tmpA).

- *Vibrio harveyi* chitobiase (gene chb).
- *Yersinia* virulence plasmid protein yscJ.
- Halocyanin from *Natrobacterium pharaonis* [4], a membrane associated copper-binding protein. This is the first archaeobacterial protein known to be modified in such a fashion).

5 From the precursor sequences of all these proteins, a consensus pattern was derived and a set of rules to identify this type of post-translational modification.

Consensus pattern {DERK}(6)-[LIVMFWSTAG](2)-[LIVMFYSTAGCQ]-[AGS]-C [C is the lipid attachment site] Additional rules: 1) The cysteine must be between positions 15 and
10 35 of the sequence in consideration. 2) There must be at least one Lys or one Arg in the first seven positions of the sequence. Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT some 100 prokaryotic proteins. Some of them are not membrane lipoproteins, but at least half of them could be.

15 [1] Hayashi S., Wu H.C. J. Bioenerg. Biomembr. 22:451-471(1990).

[2] Klein P., Somorjai R.L., Lau P.C.K. Protein Eng. 2:15-20(1988).

[3] von Heijne G. Protein Eng. 2:531-534(1989).

[4] Mattar S., Scharf B., Kent S.B.H., Rodewald K., Oesterhelt D., Engelhard M. J. Biol. Chem. 269:14939-14945(1994).

901. Ribosome recycling factor (RRF)

The ribosome recycling factor (RRF / ribosome release factor) dissociates the ribosome from the mRNA after termination of translation, and is essential bacterial growth [1]. Thus

25 ribosomes are "recycled" and ready for another round of protein synthesis. Number of members: 27

[1] Janosi L, Shimizu I, Kaji A; Medline: 94240115 Ribosome recycling factor (ribosome releasing factor) is essential for bacterial growth." Proc Natl Acad Sci U S A 1994;91:4249-
30 4253.

902. S-layer homology(SLH)

S-layers are paracrystalline mono-layered assemblies of (glyco)proteins which coat the surface of bacteria [1]. Several S-layer proteins and some other cell wall proteins contain one or more copies of a domain of about 50-60 residues, which has been called SLH (for S-layer homology) [2]. There is strong evidence that this domain serves as an anchor to the

peptidoglycan [3]. The SLH domain has been found in:

- S-layer glycoprotein of *Acetogenium kivui* (3 copies).
- S-layer 125 Kd protein of *Bacillus sphaericus* (3 copies).
- S-layer protein of *Bacillus anthracis* (3 copies).
- S-layer protein of *Bacillus licheniformis* (3 copies).
- S-layer protein (HWP) from *Bacillus brevis* strain HPD31 (3 copies).
- Middle cell wall protein (MWP) from *Bacillus brevis* strain 47 (3 copies).
- S-layer protein (p100) of *Thermus thermophilus* (1 copy).
- Outer membrane protein Omp-alpha from *Thermotoga maritima* (1 copy).
- Cellulosome anchoring protein (gene *ancA*), outer layer protein B (OlpB) and a further potential cell surface glycoprotein from *Clostridium thermocellum* (3 copies; the first copy is missing its N-terminal third which is appended to the end of the third copy; may have arisen by circular permutation).
- Amylopullulanase (gene *amyB*) from *Thermoanaerobacter thermosulfurogenes* (3 copies)
- Amylopullulanase (gene *aapT*) from *Bacillus* strain XAL-601 (3 copies).
- Endoglucanase from *Bacillus* strain KSM-635 (3 copies).
- Exoglucanase (gene *xynX*) from *Clostridium thermocellum* (3 copies).
- Xylanase A (gene *xynA*) from *Thermoanaerobacter saccharolyticum* (2 copies; 3 copies if a frameshift is taken into account).
- Protein involved in butirosin production (**ButB**) from *Bacillus circulans* (2 incomplete copies; 3 copies if three frameshifts are taken into account).
- Two hypothetical proteins from *Synechocystis* strain PCC 6803 (1 copy each).
- A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from *Bacillus circulans* (fragment of 1 copy; 3 copies if two frameshifts are taken into account).

SLH domains are found at the N- or C-termini of mature proteins. They occur in single copy followed by a predicted coiled coil domain, or in three contiguous copies. Structurally, the SLH domain is predicted to contain two alpha-helices flanking a beta strand. The SLH sequences are fairly divergent with an average identity of about 25%. It is however possible

to build a sequence pattern that starts at the second position of the domain and that spans 3/4 of its length.

Consensus pattern[LVFYT]-x-[DA]-x(2,5)-[DNGSATPHY]-[FYWPDA]-x(4)-[LIV]-x(2)-
5 [GTALV]-x(4,6)-[LIVFYC]-x(2)-G-x-[PGSTA]-x(2,3)-[MFYA]-x- [PGAV]-x(3,10)-
[LIVMA]-[STKR]-[RY]-x-[EQ]-x-[STALIVM] Sequences known to belong to this class
detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.

[1] Beveridge T.J. Curr. Opin. Struct. Biol. 4:204-212(1994).

10 [2] Lupas A., Engelhardt H., Peters J., Santarius U., Volker S., Baumeister W. J. Bacteriol.
176:1224-1233(1994).

[3] Lemaire M., Ohayon H., Gounon P., Fujino T., Beguin P. J. Bacteriol. 177:2451-
2459(1995).

15 903. Queuine tRNA-ribosyltransferase (TGT)

This is a family of queuine tRNA-ribosyltransferases EC:2.4.2.29, also known as tRNA-
guanine transglycosylase and guanine insertion enzyme. Queuine tRNA-ribosyltransferase
modifies tRNAs for asparagine, aspartic acid, histidine and tyrosine with queuine. It catalyses
20 the exchange of guanine-34 at the wobble position with 7-aminomethyl-7-deazaguanine, and
the addition of a cyclopentenediol moiety to 7-aminomethyl-7-deazaguanine-34 tRNA;
giving a hypermodified base queuine in the wobble position [1,2]. The aligned region contains
a zinc binding motif C-x-C-x2-C-x29-H, and important tRNA and 7-aminomethyl-
7deazaguanine binding residues [1]. Number of members: 27

25 [1] Romier C, Reuter K, Suck D, Ficner R; Medline: 96256303 Crystal structure of tRNA-
guanine transglycosylase: RNA modification by base exchange." EMBO J 1996;15:2850-
2857.

[2] Garcia GA, Koch KA, Chong S; Medline: 93287116 tRNA-guanine transglycosylase
30 from Escherichia coli. Overexpression, purification and quaternary structure." J Mol Biol
1993;231:489-497.

904. ThiC Family (ThiC)

5

10

[illegible]

15

20

25

30

906. UbiA prenyltransferase family signature (UbiA)

The following prenyltransferases are evolutionary related [1,2]:

- Bacterial 4-hydroxybenzoate octaprenyltransferase (gene ubiA).
 - Yeast mitochondrial para-hydroxybenzoate--polyprenyltransferase (gene COQ2).
 - Protoheme IX farnesyltransferase (heme O synthase) from yeast and mammals (gene
- 5 COX10) and from bacteria (genes cyoE or ctaB).

These proteins probably contain seven transmembrane segments. The best conserved region is located in a loop between the second and third of these segments and was used as a signature pattern.

10 Consensus pattern N-x(3)-[DE]-x(2)-[LIF]-D-x(2)-[VM]-x-R-[ST]-x(2)-R-x(4)-G Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.

15 [1] Melzer M., Heide L. Biochim. Biophys. Acta 1212:93-102(1994).

[2] Mogi T., Saiki K., Anraku Y. Mol. Microbiol. 14:391-398(1994).

907. Uncharacterized protein family UPF0044 signature (UPF0044)

20 The following uncharacterized proteins have been shown [1] to be highly similar:

- Bacillus subtilis hypothetical protein yqeI.
- Escherichia coli hypothetical protein yhbY and HI1333, the corresponding Haemophilus influenzae protein.
- Methanococcus jannaschii hypothetical protein MJ0652.

25 These are small proteins of 10 to 15 Kd. They can be picked up in the database by the following pattern. This pattern is located in the N-terminal part of these proteins.

30 Consensus pattern L-[ST]-x(3)-K-x(3)-[KR]-[SGA]-x-[GA]-H-x-L-x-P-[LIV]-x(2)-[LIV]-[GA]-x(2)-G Sequences known to belong to this class detected by the pattern ALL.

908. ATP synthase (C/AC39) subunit (vATP-synt_AC39)

This family includes the AC39 subunit from vacuolar ATP synthase Swiss:P32366 [1], and the C subunit from archaebacterial ATP synthase [2]. The family also includes subunit C

09669980 "08669980"

from the Sodium transporting ATP synthase from *Enterococcus hirae* Swiss:P43456 [3].

Number of members: 12

[1] Bauerle C, Ho MN, Lindorfer MA, Stevens TH; Medline: 93286119 The *Saccharomyces cerevisiae* VMA6 gene encodes the 36-kDa subunit of the vacuolar H(+)-ATPase membrane sector." J Biol Chem 1993;268:12749-12757.

[2] Wilms R, Freiberg C, Wegerle E, Meier I, Mayer F, Muller V; Medline: 96324968 Subunit structure and organization of the genes of the A1A0 ATPase from the Archaeon *Methanosarcina mazei* Go1." J Biol Chem 1996;271:18843-18852.

[3] Takase K, Kakinuma S, Yamato I, Konishi K, Igarashi K, Kakinuma Y; Medline: 94209269 Sequencing and characterization of the ntp gene cluster for vacuolar- type Na(+)-translocating ATPase of *Enterococcus hirae*." J Biol Chem 1994;269:11037-11044.

909. ATP synthase (E/31 kDa) subunit (vATP-synt_E)

This family includes the vacuolar ATP synthase E subunit [1], as well as the archaeobacterial ATP synthase E subunit [2]. Number of members: 24

[1] Foury F; Medline: 91009356 The 31-kDa polypeptide is an essential subunit of the vacuolar ATPase in *Saccharomyces cerevisiae*." J Biol Chem 1990;265:18554-18560.

[2] Wilms R, Freiberg C, Wegerle E, Meier I, Mayer F, Muller V; Medline: 96324968 Subunit structure and organization of the genes of the A1A0 ATPase from the Archaeon *Methanosarcina mazei* Go1." J Biol Chem 1996;271:18843-18852.

910. (WW)

The WW domain [1-4,E1] (also known as rsp5 or WWP) has been originally discovered as a short conserved region in a number of unrelated proteins, among them dystrophin, the gene responsible for Duchenne muscular dystrophy. The domain, which spans about 35 residues, is repeated up to 4 times in some proteins. It has been shown [5] to bind proteins with particular proline- motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. It appears to contain beta-strands grouped around four conserved aromatic positions; generally Trp. The name WW or WWP derives from the presence of these Trp as well as that of a conserved Pro.

It is frequently associated with other domains typical for proteins in signal transduction processes.

Proteins containing the WW domain are listed below.

5 - Dystrophin, a multidomain cytoskeletal protein. Its longest alternatively spliced form consists of an N-terminal actin-binding domain, followed by 24 spectrin-like repeats, a cysteine-rich calcium-binding domain and a C-terminal globular domain. Dystrophin form tetramers and is thought to have multiple functions including involvement in membrane stability, transduction of contractile forces to the extracellular environment and organization
10 of membrane specialization. Mutations in the dystrophin gene lead to muscular dystrophy of Duchenne or Becker type. Dystrophin contains one WW domain C-terminal of the spectrin-repeats.

- Utrophin, a dystrophin-like protein of unknown function.

15 - Vertebrate YAP protein is a substrate of an unknown serine kinase. It binds to the SH3 domain of the Yes oncoprotein via a proline-rich region. This protein appears in alternatively spliced isoforms, containing either one or two WW domains [6].

- Mouse NEDD-4 plays a role in the embryonic development and differentiation of the central nervous system. It contains 3 WW modules followed by a HECT domain. The human ortholog contains 4 WW domains, but the third WW domain is probably spliced resulting in
20 an alternate NEDD-4 protein with only 3 WW modules [3].

- Yeast RSP5 is similar to NEDD-4 in its molecular organization. It contains an N-terminal C2 domain (see <PDOC00380>), followed by a histidine-rich region, 3 WW domains and a HECT domain.

25 - Rat FE65, a transcription-factor activator expressed preferentially in liver. The activator domain is located within the N-terminal 232 residues of FE65, which also contain the WW domain.

- Yeast ESS1/PTF1, a putative peptidyl prolyl cis-trans isomerase from family ppiC (see <PDOC00840>). A related protein, dodo (gene dod) exists in Drosophila and in mammals (gene PIN1).

30 - Tobacco DB10 protein. The WW domain is located N-terminal to the region with similarity to ATP-dependent RNA helicases.

- IQGAP, a human GTPase activating protein acting on ras. It contains an N-terminal domain similar to fly muscle mp20 protein and a C-terminal ras GTPase activator domain.

0968980-101300

- Yeast pre-mRNA processing protein PRP40, *Caenorhabditis elegans* ZK1098.1 and fission yeast SpAC13C5.02 are related proteins with similarity to MYO2- type myosin, each containing two WW-domains at the N-terminus.

5 - *Caenorhabditis elegans* hypothetical protein C38D4.5, which contains one WW module, a PH domain (see <PDOC50003>) and a C-terminal phosphatidylinositol 3-kinase domain.

- Yeast hypothetical protein YFL010c.

For the sensitive detection of WW domains, a profile was developed which spans the whole homology region as well as a pattern.

10 Consensus pattern W-x(9,11)-[VFY]-[FYW]-x(6,7)-[GSTNE]-[GSTQCR]-[FYW]-x(2)-P
Sequences known to belong to this class detected by the pattern ALL. Other sequence(s)
detected in SWISS-PROT8. Sequences known to belong to this class detected by the
profileALL.

15 [1] Bork P., Sudol M. Trends Biochem. Sci. 19:531-533(1994).

[2] Andre B., Springael J.Y. Biochem. Biophys. Res. Commun. 205:1201-1205(1994).

[3] Hofmann K.O., Bucher P. FEBS Lett. 358:153-157(1995).

[4] Sudol M., Chen H.I., Bougeret C., Einbond A., Bork P. FEBS Lett. 369:67-71(1995).

[5] Chen H.I., Sudol M. Proc. Natl. Acad. Sci. U.S.A. 92:7819-7823(1995).

20 [6] Sudol M., Bork P., Einbond A., Kastury K., Druck T., Negrini M., Huebner K., Lehman
D. J. Biol. Chem. 270:14733-14741(1995).

911. Xeroderma pigmentosum (XP) [1] (XPG_1)

25 Xeroderma pigmentosum (XP) [1] is a human autosomal recessive disease, characterized by a
high incidence of sunlight-induced skin cancer. People's skin cells with this condition are
hypersensitive to ultraviolet light, due to defects in the incision step of DNA excision repair.
There are a minimum of seven genetic complementation groups involved in this pathway:
XP-A to XP-G. The defect in XP-G can be corrected by a 133 Kd nuclear protein called XPG
30 (or XPGC) [2].

XPG belongs to a family of proteins [2,3,4,5,6] that are composed of two main subsets:

- Subset 2, to which belongs mouse and human FEN-1, rad2 from fission yeast, and RAD27 from budding yeast. FEN-1 is a structure-specific endonuclease.

- Fission yeast *exo1*, a 5'->3' double-stranded DNA exonuclease that could act in a pathway that corrects mismatched base pairs.

- Yeast EXO1 (DHS1), a protein with probably the same function as *exo1*.
- Yeast DIN7.

Sequence alignment of this family of proteins reveals that similarities are largely confined to two regions. The first is located at the N-terminal extremity (N-region) and corresponds to the first 95 to 105 amino acids. The second region is internal (I-region) and found towards the C-terminus; it spans about 140 residues and contains a highly conserved core of 27 amino acids that includes a conserved pentapeptide (E-A-[DE]-A-[QS]). It is possible that the conserved acidic residues are involved in the catalytic mechanism of DNA excision repair in XPG. The amino acids linking the N- and I-regions are not conserved; indeed, they are largely absent from proteins belonging to the second subset.

Two signature patterns were developed for these proteins. The first corresponds to the central part of the N-region, the second to part of the I-region and includes the putative catalytic core pentapeptide.

Consensus pattern [VI]-[KRE]-P-x-[FYIL]-V-F-D-G-x(2)-[PIL]-x-[LVC]-K Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROT NONE.

Consensus pattern [GS]-[LIVM]-[PER]-[FYS]-[LIVM]-x-A-P-x-E-A-[DE]-[PAS]- [QS]-[CLM] Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROT/NONE.

Figure 1 consists of 11 sub-graphs, labeled (a) through (k), each showing the time course of a different physiological or behavioral parameter over a 10-minute period. The y-axis for all graphs ranges from 0 to 100. The x-axis for all graphs ranges from 0 to 10 minutes. The graphs show a general increase in the parameters during the intervention period, with some parameters showing a more pronounced increase than others.

- (a) Heart rate (b/min): Shows a steady increase from approximately 60 to 80 b/min.
- (b) Blood pressure (mmHg): Shows a steady increase from approximately 120 to 140 mmHg.
- (c) Blood flow (ml/min): Shows a steady increase from approximately 20 to 40 ml/min.
- (d) Blood flow (ml/min): Shows a steady increase from approximately 20 to 40 ml/min.
- (e) Blood flow (ml/min): Shows a steady increase from approximately 20 to 40 ml/min.
- (f) Blood flow (ml/min): Shows a steady increase from approximately 20 to 40 ml/min.
- (g) Blood flow (ml/min): Shows a steady increase from approximately 20 to 40 ml/min.
- (h) Blood flow (ml/min): Shows a steady increase from approximately 20 to 40 ml/min.
- (i) Blood flow (ml/min): Shows a steady increase from approximately 20 to 40 ml/min.
- (j) Blood flow (ml/min): Shows a steady increase from approximately 20 to 40 ml/min.
- (k) Blood flow (ml/min): Shows a steady increase from approximately 20 to 40 ml/min.

- 5

15

20

Number of members: 23

- 25

31

This family consist of various cytosolic long-chain acyl-CoA thioester hydrolases including human and rat [1,2]. The aligned region is repeated with in the sequence of human and rat

cytosolic long-chain acyl-CoA thioester hydrolases of this family. Long-chain acyl-CoA hydrolases hydrolyse palmitoyl-CoA to CoA and palmitate, they also catalyse the hydrolysis of other long chain fatty acyl-CoA thioesters. Long-chain acyl-CoA hydrolases are present in all living organisms and they may provide a mechanism for the control of lipid metabolism [1].

Number of members: 24

[1] Yamada J, Furihata T, Iida N, Watanabe T, Hosokawa M, Satoh T, Someya A, Nagaoka I, Suga T; Medline: 97236308 Molecular cloning and expression of cDNAs encoding rat brain and liver cytosolic long-chain acyl-CoA hydrolases." Biochem Biophys Res Commun 1997;232:198-203.

[2] Broustas CG, Larkins LK, Uhler MD, Hajra AK; Medline: 96209964 Molecular cloning and expression of cDNA encoding rat brain cytosolic acyl-coenzyme A thioester hydrolase." J Biol Chem 1996;271:10470-10476.

914. Agglutinin

Lectin (probable mannose binding)

Members of this family are plant lectins. Many if not all are mannose specific.

Number of members: 87

[1] Wright CS, Hester G; Medline: 97094989 The 2.0 A structure of a cross-linked complex between snowdrop lectin and a branched mannopentaose: evidence for two unique binding modes." Structure 1996;4:1339-1352.

915. (ANF_RECEPTORS)

Natriuretic peptides are hormones involved in the regulation of fluid and electrolyte homeostasis. These hormones stimulate the intracellular production of cyclic GMP as a second messenger.

Currently, three types of natriuretic peptide receptors are known [1,2]. Two express guanylate cyclase activity: GC-A (or ANP-A) which seems specific to atrial natriuretic peptide (ANP), and GC-B (or ANP-B) which seems to be stimulated more effectively by brain natriuretic

peptide (BNP) than by ANP. The third receptor (ANP-C) is probably responsible for the clearance of ANP from the circulation and does not play a role in signal transduction.

GC-A and GC-B are plasma membrane-bound proteins that share the following topology: an N-terminal extracellular domain which acts as the ligand binding region, then a transmembrane domain followed by a large cytoplasmic C-terminal region that can be subdivided into two domains: a protein kinase-like domain (see <PDOC00100>) that appears important for proper signalling and a guanylate cyclase catalytic domain (see <PDOC00425>). The topology of ANP-C is different: like GC-A and -B it possesses an extracellular ligand-binding region and a transmembrane domain, but its cytoplasmic domain is very short.

A pattern was developed from the ligand-binding region of natriuretic peptide receptors based on a highly conserved region located in the N-terminal part of the domain.

Consensus patternG-P-x-C-x-Y-x-A-A-x-V-x-R-x(3)-H-W Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROTNONE.

[1] Garbers D.L. New Biol. 2:499-504(1990).

[2] Schulz S., Chinkers M., Garbers D.L. FASEB J. 2:2026-2035(1989).

916. (Apocytochrome)

Cytochrome c family heme-binding site signature

In proteins belonging to cytochrome c family [1], the heme group is covalently attached by thioether bonds to two conserved cysteine residues. The consensus sequence for this site is Cys-X-X-Cys-His and the histidine residue is one of the two axial ligands of the heme iron. This arrangement is shared by all proteins known to belong to cytochrome c family, which presently includes cytochromes c, c', c1 to c6, c550 to c556, cc3/Hmc, cytochrome f and reaction center cytochrome c.

Consensus patternC-{CPWHF}-{CPWR}-C-H-{CFYW} Sequences known to belong to this class detected by the patternALL, except for four cytochrome c's which lack the first thioether bond. Other sequence(s) detected in SWISS-PROT454.

Note: some cytochrome c's have more than a single bound heme group c4 has 2, c7 has 3, c3 has 4, the reaction center has 4, and cc3/Hmc has 16 !

5 [1] Mathews F.S. Prog. Biophys. Mol. Biol. 45:1-56(1985).

917. ATP-synt A-c. ATP synthase Alpha chain, C terminal

[1] Medline: 94344236. Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. Abrahams JP, Leslie AG, Lutter R, Walker JE; Nature 1994;370:621-628.

10 Number of members: 125

918. (Basic)

Myc-type, 'helix-loop-helix' dimerization domain signature

HELIX_LOOP_HELIX

A number of eukaryotic proteins, which probably are sequence specific DNA-binding proteins that act as transcription factors, share a conserved domain of 40 to 50 amino acid residues. It has been proposed [1] that this domain is formed of two amphipathic helices joined by a variable length linker region that could form a loop. This 'helix-loop-helix' (HLH) domain mediates protein dimerization and has been found in the proteins listed below [2,3,E1,E2]. Most of these proteins have an extra basic region of about 15 amino acid residues that is adjacent to the HLH domain and specifically binds to DNA. They are referred as basic helix-loop-helix proteins (bHLH), and are classified in two groups: class A (ubiquitous) and class B (tissue-specific). Members of the bHLH family bind variations on the core sequence 'CANNTG', also referred to as the E-box motif. The homo- or heterodimerization mediated by the HLH domain is independent of, but necessary for DNA binding, as two basic regions are required for DNA binding activity. The HLH proteins lacking the basic domain (Emc, Id) function as negative regulators since they form heterodimers, but fail to bind DNA. The hairy-related proteins (hairy, E(spl), deadpan) also repress transcription although they can bind DNA. The proteins of this subfamily act together with co-repressor proteins, like groucho, through their C-terminal motif WRPW.

- The myc family of cellular oncogenes [4], which is currently known to contain four members: c-myc [E3], N-myc, L-myc, and B-myc. The myc genes are thought to play a role in cellular differentiation and proliferation.

5

10

15

20

25

30

- *Drosophila* extra-macrochaetae (emc) protein, which participates in sensory organ patterning by antagonizing the neurogenic activity of the achaete- scute complex. Emc is the homolog of mammalian Id proteins.

- *Drosophila* achaete-scute (AS-C) complex proteins T3 (l'sc), T4 (scute), T5 (achaete) and T8 (asense). The AS-C proteins are involved in the determination of the neuronal precursors in the peripheral nervous system and the central nervous system.

- Drosophila atonal protein (ato) which is involved in neurogenesis.

- *Drosophila* deadpan (dnp), a hairy-like protein involved in the functional differentiation of neurons.

- *Drosophila delilah* (dei) protein, which is plays an important role in the differentiation of epidermal cells into muscle.

- **Drosophila hairy (h) protein, a transcriptional repressor which regulates the embryonic segmentation and adult bristle patterning.**

- *Drosophila* enhancer of split proteins E(spl), that are hairy-like proteins active during neurogenesis. also act as transcriptional repressors.

- *Drosophila* twist (twi) protein, which is involved in the establishment of germ layers in embryos.

- Maize anthocyanin regulatory proteins R-S and LC.

- Yeast centromere-binding protein 1 (CPF1 or CBF1). This protein is involved in chromosomal segregation. It binds to a highly conserved DNA sequence, found in centromeres and in several promoters.

- Yeast INO2 and INO4 proteins.

- Yeast phosphate system positive regulatory protein PHO4 which interacts with the upstream activating sequence of several acid phosphatase genes.

- Yeast serine-rich protein TYE7 that is required for ty-mediated ADH2 expression.

[illegible]

- *Neurospora crassa* nuc-1, a protein that activates the transcription of structural genes for phosphorus acquisition.

- Fission yeast protein *esc1* which is involved in the sexual differentiation process.

5 The schematic representation of the helix-loop-helix domain is shown here:

xxxxxxxxxxxxxxxxxxxxxxxxxxxxx-----xxxxxxxxxxxxxxxxxxxxxxxxxxxxx Amphipathic
helix 1 Loop Amphipathic helix 2

10 The signature pattern that had been developed to detect this domain spans completely the second amphipathic helix.

Consensus pattern[DENSTAP]-[KR]-[LIVMAGSNT]-{FYWCPHKR}-[LIVMT]-[LIVM]-
x(2)-[STAV]-[LIVMSTACKR]-x-[VMFYH]-[LIVMTA]-{P}-{P}-[LIVMRKHQ]

15 Other sequence(s) detected in SWISS-PROT135.

[1] Murre C., McCaw P.S., Baltimore D. Cell 56:777-783(1989).

[2] Garrel J., Campuzano S. BioEssays 13:493-498(1991).

[3] Kato G.J., Dang C.V. FASEB J. 6:3065-3072(1992).

[4] Krause M., Fire A., Harrison S.W., Priess J., Weintraub H. Cell 63:907-919(1990).

[5] Riechmann V., van Cruuchten I., Sablitzky F. Nucleic Acids Res. 22:749-755(1994).

919. (Beta-lactamase)

Beta-lactamases classes -A, -C, and -D active site

25

Beta-lactamases (EC 3.5.2.6) [1,2] are enzymes which catalyze the hydrolysis of an amide bond in the beta-lactam ring of antibiotics belonging to the penicillin/cephalosporin family. Four kinds of beta-lactamase have been identified [3]. Class-B enzymes are zinc containing proteins whilst class -A, C and D enzymes are serine hydrolases. The three classes of serine beta-

30

lactamases are evolutionary related and belong to a superfamily [4] that also includes DD-peptidases and a variety of other penicillin-binding proteins (PBP's). All these proteins contain a Ser-x-x-Lys motif, where the serine is the active site residue. Although clearly homologous, the sequences of the three classes of serine beta-lactamases exhibit a large

5

10

15

20

25

- 30

Biotin is covalently attached at the active site of certain enzymes that transfer carbon dioxide from bicarbonate to organic acids to form cellular metabolites. Biotin protein ligase (BPL) is the enzyme responsible for attaching biotin to a specific lysine at the active site of biotin enzymes. Each organism probably has only one BPL. Biotin attachment is a two step reaction that results in the formation of an amide linkage between the carboxyl group of biotin and the epsilon-amino group of the modified lysine [2].

Number of members: 26

Escherichia coli biotin holoenzyme synthetase/bio repressor crystal structure delineates the biotin- and DNA-binding domains.” Proc Natl Acad Sci USA 1992;89:9257-9261.

5 [2] Chapman-Smith A, Cronan JE Jr; Medline: 10470036 The enzymatic biotinylation of
proteins: a post-translational modification of exceptional specificity.” Trends Biochem Sci
1999;24:359-363.

921. (BRCA2_repeat)

The alignment covers only the most conserved region of the repeat. Respiratory-chain NADH dehydrogenase 30 Kd subunit signature

[1] Bork P, Blomberg N, Nilges M; Medline: 96241568 Internal repeats in the BRCA2 protein sequence.” Nat Genet 1996;13:22-23.

Number of members: 63

922. (C6)

This domain of unknown function is found in the *C. elegans* protein Swiss:Q19522. It is presumed to be an extracellular domain. The C6 domain contains six conserved cysteine residues in most copies of the domain. However some copies of the domain are missing cysteine residues 1 and 3 suggesting that these form a disulphide bridge.

Number of members: 23

923. Cadherin cytoplasmic region (Cadherin C term)

Cadherins are vital in cell-cell adhesion during tissue differentiation. Cadherins are linked to the cytoskeleton by catenins. Catenins bind to the cytoplasmic tail of the cadherin. Cadherins cluster to form foci of homophilic binding units. A key determinant to the strength of the binding that it is mediated by cadherins is the juxtamembrane region of the cadherin. This region induces clustering and also binds to the protein p120ctn [1].

Number of members: 59

[2] Barth AI, Nathke IS, Nelson WJ; Medline: 97471931 Cadherins, catenins and APC protein: interplay between cytoskeletal complexes and signaling pathways." Curr Opin Cell Biol 1997;9:683-690.

[3] Braga VM, Machesky LM, Hall A, Hotchin NA; Medline: 97327766 The small GTPases Rho and Rac are required for the establishment of cadherin-dependent cell-cell contacts." J Cell Biol 1997;137:1421-1431.

Clathrin is the scaffold protein of the basket-like coat that surrounds coated vesicles. The soluble assembly unit, a triskelion, contains three heavy chains and three light chains in an extended three-legged structure. Each leg contains one heavy and one light chain. The N-terminus of the heavy chain is known as the globular domain, and is composed of seven repeats which form a beta propeller [1].

Number of members: 61

[1] ter Haar E, Musacchio A, Harrison SC, Kirchhausen T; Medline: 99043510 Atomic structure of clathrin: a beta propeller terminal domain joins an alpha zigzag linker.” Cell. 1998;95:563-573.

925. Respiratory-chain NADH dehydrogenase 30 Kd subunit signature (complex1_30Kd)

Respiratory-chain NADH dehydrogenase (EC 1.6.5.3) [1,2] (also known as complex I or NADH-ubiquinone oxidoreductase) is an oligomeric enzymatic complex located in the inner mitochondrial membrane which also seems to exist in the chloroplast and in cyanobacteria (as a NADH-plastoquinone oxidoreductase). Among the 25 to 30 polypeptide subunits of this bioenergetic enzyme complex there is one with a molecular weight of 30 Kd (in mammals) which has been found to be:

- Nuclear encoded, as a precursor form with a transit peptide in mammals, and in *Neurospora crassa*.

[illegible]

- Mitochondrial encoded in *Paramecium* (protein P1), and in the slime mold *Dictyostelium discoideum* (ORF 209).

- Chloroplast encoded in various higher plants (ORF 159). It is also present in bacteria:

- In the cyanobacteria *Synechocystis* strain PCC 6803 (gene *ndhJ*).

5 - Subunit C of *Escherichia coli* NADH-ubiquinone oxidoreductase (gene *nuoC*).

- Subunit NQO5 of *Paracoccus denitrificans* NADH-ubiquinone oxidoreductase.

This protein, in its mature form, consists of from 157 to 266 amino acid residues. The best conserved region is located in the C-terminal section and can be used as a signature pattern.

10 Consensus pattern E-R-E-x(2)-[DE]-[LIVMFY](2)-x(6)-[HK]-x(3)-[KRP]-x-[LIVM]-[LIVMYS] Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROT NONE.

15 [1] Ragan C.I. Curr. Top. Bioenerg. 15:1-36(1987).

[2] Weiss H., Friedrich T., Hofhaus G., Preis D. Eur. J. Biochem. 197:563-576(1991).

926. Respiratory-chain NADH dehydrogenase 49 Kd subunit signature (complex1_49Kd)

20 Respiratory-chain NADH dehydrogenase (EC 1.6.5.3) [1,2] (also known as complex I or NADH-ubiquinone oxidoreductase) is an oligomeric enzymatic complex located in the inner mitochondrial membrane which also seems to exist in the chloroplast and in cyanobacteria (as a NADH-plastoquinone oxidoreductase). Among the 25 to 30 polypeptide subunits of this bioenergetic enzyme complex there is one with a molecular weight of 49 Kd (in mammals), which is the third largest subunit of complex I and is a component of the iron-sulfur (IP) fragment of the enzyme. It seems to bind a 4Fe-4S iron-sulfur cluster. The 49 Kd subunit has been found to be:

- Nuclear encoded, as a precursor form with a transit peptide in mammals, and in *Neurospora crassa*.

30 - Mitochondrial encoded in protozoan such as *Paramecium* (ORF 400), *Leishmania* and *Trypanosoma* (MURF 3).

- Chloroplast encoded in various higher plants (ORF 392).

The 49 Kd subunit is highly similar to [3,4]:

- Subunit D of *Escherichia coli* NADH-ubiquinone oxidoreductase (gene *nuoD*).

0969990-101300

- 5

Consensus pattern [LIVMH]-H-[RT]-[GA]-x-E-K-[LIVMTN]-x-E-x-[KRQ] Sequences known to belong to this class detected by the patternALL.

- 10

15

20

25

It has been shown [3,4] that nitrous oxide reductase (EC 1.7.99.6) (gene *nosZ*) of *Pseudomonas* has sequence similarity in its C-terminus to CO II. This enzyme is part of the bacterial respiratory system which is activated under anaerobic conditions in the presence of

nitrate or nitrous oxide. NosZ is a periplasmic homodimer that contains a dinuclear copper center, probably located in a 3- dimensional fold similar to the cupredoxin-like fold that has been suggested for the copper-binding site of CO II [3].

- 5 The dinuclear purple copper center is formed by 2 histidines and 2 cysteines [5]. This region was used as a signature pattern. The conserved valine and the conserved methionine are said to be involved in stabilizing the copper-binding fold by interacting with each other.

Consensus pattern V-x-H-x(33,40)-C-x(3)-C-x(3)-H-x(2)-M [The two C's and two H's are
10 copper ligands] Sequences known to belong to this class detected by the patternALL, except for *Paramecium primaurelia* as well as in some plants where the pattern ends with Thr; an RNA editing event at this position could change this Thr to Met.

Note: cytochrome cbb(3) subunit 2 does not belong to this family.

[1] Capaldi R.A., Malatesta F., Darley-USmar V.M. *Biochim. Biophys. Acta* 726:135-148(1983).

[2] Garcia-Horsman J.A., Barquera B., Rumbley J., Ma J., Gennis R.B. *J. Bacteriol.* 176:5587-5600(1994).

[3] van der Oost J., Lappalainen P., Musacchio A., Warne A., Lemieux L., Rumbley J., Gennis R.B., Aasa R., Pascher T., Malmstrom B.G., Saraste M. *EMBO J.* 11:3209-3217(1992).

[4] Zumft W.G., Dreutsch A., Loechele S., Cuypers H., Friedrich B., Schneider B. *Eur. J. Biochem.* 208:31-40(1992).

928. Cytochrome C assembly protein (CytC_asm)

This family consists of various proteins involved in cytochrome c assembly from mitochondria and bacteria; CycK from *Rhizobium*[3], CcmC from *E. coli* and *Paracoccus denitrificans* [2,1] and orf240 from wheat mitochondria [4]. The members of this family are probably integral membrane proteins with six predicted transmembrane helices. It has been proposed that members of this family comprise a membrane component of an ABC (ATP binding cassette) transporter complex. It is also proposed that this transporter is necessary for transport of some component needed for cytochrome c assembly. One member CycK



- 5

930. Cytochrome b/b6 signatures (Cytochrome_b)

10

15

25

35

+---Fe-b562----+ | +---Fe-b566--|-+ |||

```
xxxxxxxxxxxxHxHxxxxxxxxxxxxHxHxxxxxxxxxxxxPEWxxxxxxxxxxxxxxxxxxxxx <-----
```


[1] Barten R, Meyer TF; Medline: 98273626 Cloning and characterisation of the *Neisseria gonorrhoeae* *aroB* gene." Mol Gen Genet 1998;258:34-44.

[2] Hawkins AR, Lamb HK; Medline: 96048023 The molecular biology of multidomain proteins. Selected examples." Eur J Biochem 1995;232:7-18.

The 3-dehydroquinate synthase EC:4.6.1.3 domain is present in isolation in various bacterial 3-dehydroquinate synthases and also present as a domain in the pentafunctional AROM polypeptide Swiss:P07547 [2]. 3-dehydroquinate (DHQ) synthase catalyses the formation of dehydroquinate (DHQ) and orthophosphate from 3-deoxy-D-arabino heptulosonic 7 phosphate [1]. This reaction is part of the shikimate pathway which is involved in the biosynthesis of aromatic amino acids.

Number of members: 25

933. Dihydrofolate reductase signature (DiHfolate_red)

Dihydrofolate reductases (EC 1.5.1.3) [1] are ubiquitous enzymes which catalyze the reduction of folic acid into tetrahydrofolic acid. They can be inhibited by a number of antagonists such as trimethoprim and methotrexate which are used as antibacterial or anticancerous agents. A signature pattern was derived from a region in the N-terminal part of these enzymes, which includes a conserved Pro-Trp dipeptide; the tryptophan has been shown [2] to be involved in the binding of substrate by the enzyme.

Consensus pattern[LVAGC]-[LIF]-G-x(4)-[LIVMF]-P-W-x(4,5)-[DE]-x(3)-[FYIV]-x(3)-[STIQ] Sequences known to belong to this class detected by the patternALL, except for type II bacterial, plasmid-encoded, dihydrofolate reductases which do not belong to the same class of enzymes.

[1] Harpers' Review of Biochemistry, Lange, Los Altos (1985).

[2] Bolin J.T., Filman D.J., Matthews D.A., Hamlin R.C., Kraut J. J. Biol. Chem. 257:13650-13662(1982).

934. (DIL)

Number of members: 31

DNA topoisomerase II signature (cross-reference = TOPOISOMERASE_II)

There are many regions of sequence homology between the different subtypes of topoisomerase II. The relation between the different subunits is shown in the following representation:

[-----Protein 39-*-----][----Protein 52----] Phage T4

[-----gyrB-----*-----][-----gyrA-----] Prokaryote II

[-----parE-----*-----][-----parD-----] Prokaryote IV

[-----*-----] Eukaryote and ASF

'*': Position of the pattern.

As a signature pattern for this family of proteins, a region was selected that contains a highly conserved pentapeptide. The pattern is located in *gyrB*, in *parE*, and in protein 39 of phage T4 topoisomerase.

As a signature pattern for this family of proteins, a region was selected that contains a highly conserved pentapeptide. The pattern is located in gyrB, in parE, and in protein 39 of phage T4 topoisomerase.

- 5 Consensus pattern [LIVMA]-x-E-G-[DN]-S-A-x-[STAG] Sequences known to belong to this class detected by the patternALL.

- [1] Sternglanz R. Curr. Opin. Cell Biol. 1:533-535(1990).
 [2] Bjornsti M.-A. Curr. Opin. Struct. Biol. 1:99-103(1991).
 10 [3] Sharma A., Mondragon A. Curr. Opin. Struct. Biol. 5:39-47(1995).
 [4] Roca J. Trends Biochem. Sci. 20:156-160(1995).

937. Prolyl oligopeptidase family serine active site (DPPIV_N_term)

- 15 The prolyl oligopeptidase family [1,2,3] consist of a number of evolutionary related peptidases whose catalytic activity seems to be provided by a charge relay system similar to that of the trypsin family of serine proteases, but which evolved by independent convergent evolution. The known members of this family are listed below.

- 20 - Prolyl endopeptidase (EC 3.4.21.26) (PE) (also called post-proline cleaving enzyme). PE is an enzyme that cleaves peptide bonds on the C-terminal side of prolyl residues. The sequence of PE has been obtained from a mammalian species (pig) and from bacteria (Flavobacterium meningosepticum and Aeromonas hydrophila); there is a high degree of sequence conservation between these sequences.

- 25 - Escherichia coli protease II (EC 3.4.21.83) (oligopeptidase B) (gene prtB) which cleaves peptide bonds on the C-terminal side of lysyl and arginyl residues.

- Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV). DPP IV is an enzyme that removes N-terminal dipeptides sequentially from polypeptides having unsubstituted N-termini provided that the penultimate residue is proline.

- 30 - Yeast vacuolar dipeptidyl aminopeptidase A (DPAP A) (gene: STE13) which is responsible for the proteolytic maturation of the alpha-factor precursor.

- Yeast vacuolar dipeptidyl aminopeptidase B (DPAP B) (gene: DAP2).

- Acylamino-acid-releasing enzyme (EC 3.4.19.1) (acyl-peptide hydrolase). This enzyme catalyzes the hydrolysis of the amino-terminal peptide bond of an N-acetylated protein to generate a N-acetylated amino acid and a protein with a free amino-terminus.

A conserved serine residue has experimentally been shown (in E.coli protease II as well as in pig and bacterial PE) to be necessary for the catalytic mechanism. This serine, which is part of the catalytic triad (Ser, His, Asp), is generally located about 150 residues away from the C-terminal extremity of these enzymes (which are all proteins that contains about 700 to 800 amino acids).

Consensus pattern D-x(3)-A-x(3)-[LIVMFYW]-x(14)-G-x-S-x-G-G-[LIVMFYW](2) [S is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for yeast DPAP A.

Note: these proteins belong to families S9A/S9B/S9C in the classification of peptidases [4,E1].

- [1] Rawlings N.D., Polgar L., Barrett A.J. Biochem. J. 279:907-911(1991).
- [2] Barrett A.J., Rawlings N.D. Biol. Chem. Hoppe-Seyler 373:353-360(1992).
- [3] Polgar L., Szabo E. Biol. Chem. Hoppe-Seyler 373:361-366(1992).
- [4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).

938. Deoxyhypusine synthase (DS)

Eukaryotic initiation factor 5A (eIF-5A) contains an unusual amino acid, hypusine [N epsilon-(4-aminobutyl-2-hydroxy)lysine]. The first step in the post-translational formation of hypusine is catalysed by the enzyme deoxyhypusine synthase (DS) EC:1.1.1.249. The modified version of eIF-5A, and DS, are required for eukaryotic cell proliferation [1].

Number of members: 9

[1] Liao DI, Wolff EC, Park MH, Davies DR; Medline: 98154315 Crystal structure of the NAD complex of human deoxyhypusine synthase: an enzyme with a ball-and-chain mechanism for blocking the active site." Structure 1998;6:23-32.

939. (DUF21)

Many of the sequences in this family are annotated as hemolysins, however this is due to a similarity to Swiss:Q54318 that does not contain this domain. This domain is found in the N-terminus of the proteins adjacent to two intracellular CBS domains CBS.

Number of members: 42

940. (DUF59)

This family includes prokaryotic proteins of unknown function. The family also includes PhaH Swiss:O84984 from *Pseudomonas putida*. PhaH forms a complex with PhaF Swiss:O84982, PhaG Swiss:O84983 and PhaI Swiss:O84985, which hydroxylates phenylacetic acid to 2-hydroxyphenylacetic acid [1]. So members of this family may all be components of ring hydroxylating complexes.

Number of members: 15

[1] Olivera ER, Minambres B, Garcia B, Muniz C, Moreno MA, Ferrandez A, Diaz E, Garcia JL, Luengo JM; Medline: 98263372 Molecular characterization of the phenylacetic acid catabolic pathway in *Pseudomonas putida* U: the phenylacetyl-CoA catabolon." Proc Natl Acad Sci U S A 1998;95:6419-6424.

941. (DUF82)

The protein contains four conserved cysteines that may be involved in metal binding or disulphide bridges.

Number of members: 4

942. Riboflavin kinase / FAD synthetase (FAD_Synth)

This family consists part of the bifunctional enzyme riboflavin kinase / FAD synthetase. These enzymes have both ATP:riboflavin 5'-phospho transferase and ATP:FMN-adenylyltransferase activities [1]. They catalyse the 5'-phosphorylation of riboflavin to FMN and the adenylation of FMN to FAD [1].

- 10 [1] Romao MJ, Archer M, Moura I, Moura JJ, LeGall J, Engh R, Schneider M, Hof P, Huber
R; Medline: 96072968 Crystal structure of the xanthine oxidase-related aldehyde oxido-
reductase from *D. gigas*.” Science 1995;270:1170-1176.

Number of members: 53

- 15 944. Filovirus glycoprotein (Filo_glycop)

20 function [1]. Processing of this protein may be involved in viral pathogenicity [2].

[1] Volchkov VE, Feldmann H, Volchkova VA, Klenk HD; Medline: 98245155 Processing of the Ebola virus glycoprotein by the proprotein convertase furin." Proc Natl Acad Sci U S A 1998;95:5762-5767.

- 25 A 1998;95:5762-5767.

- 30 945. Frataxin-like domain (Frataxin_Cyay)

This family contains proteins that have a domain related to the globular C-terminus of Frataxin the protein that is mutated in Friedreich's ataxia. This domain is found in a family of bacterial proteins. The function of this domain is currently unknown.

Number of members: 12

[1] Gibson TJ, Koonin EV, Musco G, Pastore A, Bork P; Medline: 97084946 Friedrich's ataxia protein: phylogenetic evidence for mitochondrial dysfunction." Trends Neurosci 1996;19:465-468.

946. (GAF)

Domain present in phytochromes and cGMP-specific phosphodiesterases.

Number of members: 296

[1] Aravind L, Ponting CP; Medline: 98094688 The GAF domain: an evolutionary link between diverse phototransducing proteins." Trends Biochem Sci 1997;22:458-459.

947. Galaptin signature (Gal-bind_lectin)

All vertebrates synthesize soluble galactoside-binding lectins [1,2,3] (also known as galectins, galaptins or S-lectin). These carbohydrate-binding proteins are developmentally regulated. Although their exact physiological role is not yet clear they seem to be involved in differentiation, cellular regulation and tissue construction. The sequence of galactoside-binding lectins from electric eel (electrolectin), conger eel (congerin), chicken and a number of mammalian species is known. These lectins are proteins of about 130 to 140 amino acid residues (14 Kd to 16 Kd).

A number of other proteins are known to belong to this family:

- Galectin-3 (also known as MAC-2 antigen; CBP-35 or IgE-binding protein), a 35 Kd lectin which binds immunoglobulin E and which is composed of two domains: a N-terminal domain that consist of tandem repeats of a glycine/ proline-rich sequence and a C-terminal galaptin domain.
- Galectin-4 [4], which is composed of two galaptin domains.
- Galectin-5.
- Galectin-7 [5], a keratinocyte protein which could be involved in cell-cell and/or cell-matrix interactions necessary for normal growth control.
- Galectin-8 [6], which is composed of two galaptin domains.

0965990-10100

5

inflammation.

- *Caenorhabditis elegans* *lec-7* and *lec-8*.

10

15

- 20

25

30

GLYCOSYL_HYDROL_F2_1; PS00608; GLYCOSYL_HYDROL_F2_2

It has been shown [1,2,E1] that the following glycosyl hydrolases can be, on the basis of sequence similarities, classified into a single family:

-Beta-galactosidases (EC 3.2.1.23) from bacteria such as *Escherichia coli* (genes *lacZ* and *ebgA*), *Clostridium acetobutylicum*, *Clostridium thermosulfurogenes*, *Klebsiella pneumoniae*, *Lactobacillus delbrueckii*, or *Streptococcus thermophilus* and from the fungi *Kluyveromyces lactis*.

-Beta-glucuronidase (EC 3.2.1.31) from *Escherichia coli* (gene *uidA*) and from mammals. One of the conserved regions in these enzymes is centered on a conserved glutamic acid residue which has been shown [3], in *Escherichia coli lacZ*, to be the general acid/base catalyst in the active site of the enzyme. This region has been used as a signature pattern. A highly conserved region located some sixty residues upstream from the active site glutamate has been selected as a second signature pattern.

Consensus pattern N-x-[LIVMFYWD]-R-[STACN](2)-H-Y-P-x(4)-[LIVMFYWS](2)-x(3)-[DN]-x(2)-G-[LIVMFYW](4) Sequences known to belong to this class detected by the pattern ALL.

Consensus pattern [DENQLF]-[KRVW]-N-[HRY]-[STAPPV]-[SAC]-[LIVMFS](3)-W-[GS]-x(2,3)-N-E [E is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for *Rhizobium meliloti lacZ*.

[1]Henrissat B. *Biochem. J.* 280:309-316(1991).

[2]Schroeder C.J., Robert C., Lenzen G., McKay L.L., Mercenier A. J. *Gen. Microbiol.* 137:369-380(1991).

[3]Gebler J.C., Aebersold R., Withers S.G. *J. Biol. Chem.* 267:11126-11130(1992).

952. (Glyco_hydro_3) Glycosyl hydrolases family 3 active site

PROSITE: PDOC00621. PROSITE cross-reference(s)PS00775; GLYCOSYL_HYDROL_F3

It has been shown [1,2] that the following glycosyl hydrolases can be, on the basis of sequence similarities, classified into a single family:

-Beta glucosidases (EC 3.2.1.21) from the fungi *Aspergillus wentii* (A-3), *Hansenula anomala*, *Kluyveromyces fragilis*, *Saccharomycopsis fibuligera*, (BGL1 and BGL2), *Schizophyllum commune* and *Trichoderma reesei* (BGL1).

-Beta glucosidases from the bacteria *Agrobacterium tumefaciens* (Cbg1), *Butyrivibrio fibrisolvens* (bglA), *Clostridium thermocellum* (bglB), *Escherichia coli* (bglX), *Erwinia chrysanthemi* (bgxA) and *Ruminococcus albus*.

-*Alteromonas* strain O-7 beta-hexosaminidase A (EC 3.2.1.52).

5 -*Bacillus subtilis* hypothetical protein yzbA.

-*Escherichia coli* hypothetical protein ycfO and HI0959, the corresponding *Haemophilus influenzae* protein.

One of the conserved regions in these enzymes is centered on a conserved aspartic acid residue which has been shown [3], in *Aspergillus wentii* beta-glucosidase A3, to be implicated in the catalytic mechanism. This region was used as a signature pattern.

Consensus pattern[LIVM](2)-[KR]-x-[EQK]-x(4)-G-[LIVMFT]-[LIVT]-[LIVMF]-[ST]-D-x(2)-[SGADNI] [D is the active site residue]

Sequences known to belong to this class detected by the patternALL.

[1]Henrissat B. *Biochem. J.* 280:309-316(1991).

[2]Castle L.A., Smith K.D., Morris R.O. *J. Bacteriol.* 174:1478-1486(1992).

[3]Bause E., Legler G. *Biochim. Biophys. Acta* 626:459-465(1980).

953. GP120 - Envelope glycoprotein GP120

The entry of HIV requires interaction of viral GP120 with Swiss:P01730 and a chemokine receptor on the cell surface. Number of members: 17891

[1]Medline: 98303379. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA; *Nature* 1998;393:648-659.

954. (GSPII_E) Bacterial type II secretion system protein E signature

PROSITE: PDOC00567. PROSITE cross-reference(s) PS00662; T2SP_E

A number of bacterial proteins, some of which are involved in a general secretion pathway (GSP) for the export of proteins (also called the type II pathway) [1,2], have been found to be evolutionary related. These proteins are listed below:

-The 'E' protein from the GSP operon of: *Aeromonas* (gene *exeE*); *Erwinia* (gene *outE*); *Escherichia coli* (gene *yheG*); *Klebsiella pneumoniae* (gene *pulE*); *Pseudomonas aeruginosa* (gene *xcpR*); *Vibrio cholerae* (gene *epsE*) and *Xanthomonas campestris* (gene *xpsE*).

-*Agrobacterium tumefaciens* Ti plasmid *virB* operon protein 11. This protein is required for the transfer of T-DNA to plants.

-*Bacillus subtilis* *comG* operon protein 1 which is required for the uptake of DNA by competent *Bacillus subtilis* cells.

-*Aeromonas hydrophila* *tapB*, involved in type IV pilus assembly.

-*Pseudomonas* protein *pilB*, which is essential for the formation of the pili.

-*Pseudomonas aeruginosa* protein twitching mobility protein *pilT*.

-*Neisseria gonorrhoeae* type IV pilus assembly protein *pilF*.

-*Vibrio cholerae* protein *tcpT*, which is involved in the biosynthesis of the *tcp* pilus.

-*Escherichia coli* protein *hofB* (*hopB*).

-*Escherichia coli* hypothetical protein *ygcB*.

-*Escherichia coli* hypothetical protein *yggR*.

These proteins have from 344 (*pilT* and *virB11*) to 568 (*tapB*) amino acids, they are probably cytoplasmically located and, on the basis of the presence of a conserved P-loop region (see <PDOC00017>), probably bind ATP. A region that overlaps the 'B' motif of ATP-binding proteins was selected as a signature pattern.

Consensus pattern[LIVM]-R-x(2)-P-D-x-[LIVM](3)-G-E-[LIVM]-R-D

Sequences known to belong to this class detected by the patternALL, except for *ygcB*.

[1]Salmond G.P.C., Reeves P.J. Trends Biochem. Sci. 18:7-12(1993).

[2]Hobbs M., Mattick J.S. Mol. Microbiol. 10:233-243(1993).

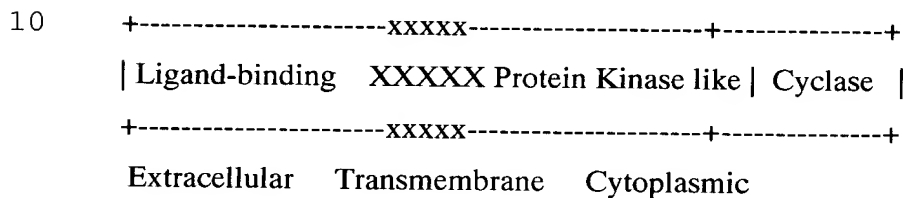
955. (guanylate_cyc) Guanylate cyclases signature

PROSITE: PDOC00425. PROSITE cross-reference(s) PS00452;

GUANYLATE_CYCLASES Guanylate cyclases (EC 4.6.1.2) [1 to 4] catalyze the formation of cyclic GMP (cGMP) from GTP. cGMP acts as an intracellular messenger, activating cGMP dependent kinases and regulating CGMP-sensitive ion channels. The role of cGMP as a second messenger in vascular smooth muscle relaxation and retinal photo-transduction is well established. Guanylate cyclase is found both in the soluble and particular

fraction of eukaryotic cells. The soluble and plasma membrane-bound forms differ in structure, regulation and other properties.

Most currently known plasma membrane-bound forms are receptors for small polypeptides. The topology of such proteins is the following: they have a N-terminal extracellular domain which acts as the ligand binding region, then a transmembrane domain, followed by a large cytoplasmic C-terminal region that can be subdivided into two domains: a protein kinase-like domain that appears important for proper signalling and a cyclase catalytic domain. This topology is schematically represented below.



15 The known guanylate cyclase receptors are:

- The sea-urchins receptors for speract and resact, which are small peptides that stimulate sperm motility and metabolism.
- The receptors for natriuretic peptides (ANF). Two forms of ANF receptors with guanylate cyclase activity are currently known: GC-A (or ANP-A) which seems specific to atrial natriuretic peptide (ANP), and GC-B (or ANP-B) which seems to be stimulated more effectively by brain natriuretic peptide (BNP) than by ANP.
- The receptor for Escherichia coli heat-stable enterotoxin (GC-C). The endogenous ligand for this intestinal receptor seems to be a small peptide called guanylin.
- Retinal guanylate cyclase (retGC) which probably plays a specific functional role in the rods and/or cones of photoreceptors. It is not known if this protein acts as receptor, but its structure is similar to that of the other plasma membrane-bound GCs.

25 The soluble forms of guanylate cyclase are cytoplasmic heterodimers. The two subunits, alpha and beta are proteins of from 70 to 82 Kd which are highly related. Two forms of beta subunits are currently known: beta-1 which seems to be expressed in lung and brain, and beta-2 which is more abundant in kidney and liver.

30 The membrane and cytoplasmic forms of guanylate cyclase share a conserved domain which is probably important for the catalytic activity of the enzyme. Such a domain is also found twice in the different forms of membrane-bound adenylate cyclases (also known as

class-III) [5,6] from mammals, slime mold or *Drosophila*. A consensus pattern was derived from the most conserved region in that domain.

Consensus pattern G-V-[LIVM]-x(0,1)-G-x(5)-[FY]-x-[LIVM]-[FYW]-[GS]-[DNTHKW]-
5 [DNT]-[IV]-[DNTA]-x(5)-[DE]

Sequences known to belong to this class detected by the pattern ALL, except for the sea urchin *Arbacia punctulata* resact receptor which lack this domain.

Note this pattern will detect both domains of adenylate cyclases class-III.

10 [1]Koesling D., Boehme E., Schultz G. FASEB J. 5:2785-2791(1991).

[2]Garbers D.L. New Biol. 2:499-504(1990).

[3]Garbers D.L. Cell 71:1-4(1992).

[4]Yuen P.S.T., Garbers D.L. Annu. Rev. Neurosci. 15:193-225(1992).

[5]Iyengar R. FASEB J. 7:768-775(1993).

15 [6]Barzu O., Danchin A. Prog. Nucleic Acid Res. Mol. Biol. 49:241-283(1994).

956. Hemolysin-type calcium-binding region signature (HemolysinCabinD)

Gram-negative bacteria produce a number of proteins which are secreted into the growth
20 medium by a mechanism that does not require a cleaved N-terminal signal sequence. These proteins, while having different functions, seem [1] to share two properties: they bind calcium and they contain a variable number of tandem repeats consisting of a nine amino acid motif rich in glycine, aspartic acid and asparagine. It has been shown [2] that such a domain is involved in the binding of calcium ions in a parallel beta roll structure. The proteins which
25 are currently known to belong to this category are:

- Hemolysins from various species of bacteria. Bacterial hemolysins are exotoxins that attack blood cell membranes and cause cell rupture. The hemolysins which are known to contain such a domain are those from: *E. coli* (gene hlyA), *A. pleuropneumoniae* (gene appA), *A. actinomycetemcomitans* and *P. haemolytica* (leukotoxin) (gene lktA).

- 30 - Cyclolysin from *Bordetella pertussis* (gene cyaA). A multifunctional protein which is both an adenylate cyclase and a hemolysin.

- Extracellular zinc proteases: serralysin (EC 3.4.24.40) from *Serratia*, prtB and prtC from *Erwinia chrysanthemi* and aprA from *Pseudomonas aeruginosa*.

- Nodulation protein nodO from *Rhizobium leguminosarum*.

09689980 "101300

A signature pattern was derived from conserved positions in the sequence of the calcium-binding domain.

Consensus pattern D-x-[LI]-x(4)-G-x-D-x-[LI]-x-G-G-x(3)-D Sequences known to belong to this class detected by the pattern ALL.

Note: This pattern is found once in nodO and the extracellular proteases but up to 5 times in some hemolysin/cyclolysins.

[1] Economou A., Hamilton W.D.O., Johnston A.W.B., Downie J.A. EMBO J. 9:349-354(1990).

[2] Baumann U., Wu S., Flaherty K.M., McKay D.B. EMBO J. 12:3357-3364(1993).

957. Hint module (Hint)

This is an alignment of the Hint module in the Hedgehog proteins. It does not include any Inteins which also possess the Hint module.

Number of members: 36

[1] Hall TM, Porter JA, Young KE, Koonin EV, Beachy PA, Leahy DJ; Medline: 97474313
Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins." Cell 1997;91:85-97.

958. Hydantoinase/oxoprolinase (Hydantoinase)

This family includes the enzymes hydantoinase and oxoprolinase EC:3.5.2.9. Both reactions involve the hydrolysis of 5-membered rings via hydrolysis of their internal imide bonds [1].

Number of members: 14

[1] Ye GJ, Breslow EB, Meister A, Guo-jie GE\$[corrected to Ye GJ]; Medline: 97113037
The amino acid sequence of rat kidney 5-oxo-L-prolinase determined by cDNA cloning"
[published erratum appears in J Biol Chem 1997 Feb 14;272(7):4646] J Biol Chem
1996;271:32293-32300.

959. IMP dehydrogenase / GMP reductase signature (IMPDH_N)

IMP dehydrogenase (EC 1.1.1.205) (IMPDH) catalyzes the rate-limiting reaction of de novo GTP biosynthesis, the NAD-dependent reduction of IMP into XMP [1]. Inhibition of IMP dehydrogenase activity results in the cessation of DNA synthesis. As IMP dehydrogenase is associated with cell proliferation, it is a possible target for cancer chemotherapy. Mammalian and bacterial IMPDHs are tetramers of identical chains. There are two IMP dehydrogenase isozymes in humans [2].

GMP reductase (EC 1.6.6.8) catalyzes the irreversible and NADPH-dependent reductive deamination of GMP into IMP [3]. It converts nucleobase, nucleoside and nucleotide derivatives of G to A nucleotides, and maintains intracellular balance of A and G nucleotides.

IMP dehydrogenase and GMP reductase share many regions of sequence similarity. One of these regions is centered on a cysteine residue thought [3] to be involved in binding IMP. This region was used as a signature pattern.

Consensus pattern[LIVM]-[RK]-[LIVM]-G-[LIVM]-G-x-G-S-[LIVM]-C-x-T [C is the putative IMP-binding residue] Sequences known to belong to this class detected by the pattern ALL.

[1] Collart F.R., Huberman E. J. Biol. Chem. 263:15769-15772(1988).

[2] Natsumeda Y., Ohno S., Kawasaki H., Konno Y., Weber G., Suzuki K. J. Biol. Chem. 265:5292-5295(1990).

[3] Andrews S.C., Guest J.R. Biochem. J. 255:35-43(1988).

960. impB/mucB/samB family (IMS)

These proteins are involved in UV protection (Swiss).

Number of members: 38

961. Type II intron maturase (Intron_maturas2)

096658960

Group II introns use intron-encoded reverse transcriptase, maturase and DNA endonuclease activities for site-specific insertion into DNA [2]. Although this type of intron is self splicing in vitro they require a maturase protein for

splicing in vivo. It has been shown that a specific region of the aI2 intron is needed for the maturase function [1]. This region was found to be conserved in group II introns and called domain X [3].

Number of members: 335

[1] Moran JV, Mecklenburg KL, Sass P, Belcher SM, Mahnke D, Lewin A, Perlman P; Medline: 94301788 Splicing defective mutants of the COXI gene of yeast mitochondrial DNA: initial definition of the maturase domain of the group II intron aI2. Nucleic Acids Res 1994;22:2057-2064.

[2] Guo H, Zimmerly S, Perlman PS, Lambowitz AM; Medline: 98031910 Group II intron endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA." EMBO J 1997;16:6835-6848.

[3] Mohr G, Perlman PS, Lambowitz AM; Medline: 94077696 Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function." Nucleic Acids Res 1993;21:4991-4997.

962. LAGLIDADG endonuclease (Intron_maturase)

[1] Heath PJ, Stephens KM, Monnat RJ Jr, Stoddard BL; Medline: 97331323 The structure of I-Crel, a group I intron-encoded homing endonuclease." Nat Struct Biol 1997;4:468-476.

[2] Belfort M, Roberts RJ; Medline: 97402526 Homing endonucleases: keeping the house in order." Nucleic Acids Res 1997;25:3379-3388.

[3] Dalgaard JZ, Klar AJ, Moser MJ, Holley WR, Chatterjee A, Mian IS; Medline: 98026854 Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family." Nucleic Acids Res 1997;25:4626-4638.

Number of members: 220

963. Isopentenyl transferase (IPT)

the binding of calcium and manganese; the second one is located in the N-terminal of the alpha chain.

Consensus pattern [LIV]-[STAG]-V-[DEQV]-[FLI]-D-[ST] [D binds manganese and calcium] Sequences known to belong to this class detected by the pattern ALL.

Consensus pattern [LIV]-x-[EDQ]-[FYWKR]-V-x-[LIVF]-G-[LF]-[ST] Sequences known to belong to this class detected by the pattern ALL.

[1] Sharon N., Lis H. FASEB J. 4:3198-320(1990).

[2] Lis H., Sharon N. Annu. Rev. Biochem. 55:33-37(1986).

966. Malate synthase signature (malate_synthase)

Malate synthase (EC 4.1.3.2) catalyzes the aldol condensation of glyoxylate with acetyl-CoA to form malate - the second step of the glyoxylate bypass, an alternative to the tricarboxylic acid cycle in bacteria, fungi and plants. Malate synthase is a protein of 530 to 570 amino acids whose sequence is highly conserved across species [1]. As a signature pattern, a very conserved region was selected in the central section of the enzyme.

Consensus pattern[KR]-[DENQ]-H-x(2)-G-L-N-x-G-x-W-D-Y-[LIVM]-F Sequences known to belong to this class detected by the pattern ALL.

[1] Bruinenberg P.G., Blaauw M., Kazemier B., Ab G. Yeast 6:245-254(1990).

967. MatK/TrnK amino terminal region (MatK_N)

[1] Mohr G, Perlman PS, Lambowitz AM; Medline: 94077696 Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function." Nucleic Acids Res 1993;21:4991-4997.

Number of members: 495

968. MOZ/SAS family (MOZ_SAS)

This region of these proteins has been suggested to be homologous to acetyltransferases [1]. However the similarity is not supported by standard sequence analysis.

Number of members: 15

5

[1] Kamine J, Elangovan B, Subramanian T, Coleman D, Chinnadurai G; Medline: 96182937 Identification of a cellular protein that specifically interacts with the essential cysteine region of the HIV-1 Tat transactivator." Virology 1996;216:357-366.

10

[2] Reifsnyder C, Lowell J, Clarke A, Pillus L; Medline: 96376969 Yeast SAS silencing genes and human genes associated with AML and HIV-1 Tat interactions are homologous with acetyltransferases" [see comments] [published erratum appears in Nat Genet 1997 May;16(1):109] Nat Genet 1996;14:42-49.

969. mRNA capping enzyme (mRNA_cap_enzyme)

15

[1] Hakansson K, Doherty AJ, Shuman S, Wigley DB; Medline: 97304383 X-ray crystallography reveals a large conformational change during guanylyl transfer by mRNA capping enzymes." Cell 1997;89:545-553.

20

Number of members: 7

970. DNA mismatch repair proteins mutS family signature (MutS_C)

25

Mismatch repair contributes to the overall fidelity of DNA replication [1]. It involves the correction of mismatched base pairs that have been missed by the proofreading element of the DNA polymerase complex. The sequence of some proteins involved in mismatch repair in different organisms have been found to be evolutionary related [2,3]. One of these families is called mutS [4,E1], it consists of:

30

- Prokaryotic protein mutS protein (also called hexA in *Streptococcus pneumoniae*). Muts is thought to carry out the mismatch recognition step of DNA repair.
- Eukaryotic MSH1, which is involved in mitochondrial DNA repair.
- Eukaryotic MSH2, which is involved in nuclear postreplication mismatch repair. MSH2 heterodimerizes with MSH6. In man, MSH2 is involved in a form of familial hereditary nonpolyposis colon cancer (HNPCC).

- Eukaryotic MSH3, which is probably involved in the repair of large loops.
- Eukaryotic MSH4, which is involved in meiotic recombination.
- Eukaryotic MSH5, which is involved in meiotic recombination.
- Eukaryotic MSH6 (also known as G/T mismatch binding protein), a DNA-repair protein that binds to G/T mismatches through heterodimerization with MSH2.
- Prokaryotic protein mutS2 whose function is not yet known.
- A coral (*Sarcophyton glaucum*) mitochondrial encoded mutS-like protein.

As a signature pattern for this class of mismatch repair proteins a region rich in glycine and negatively charged residues was selected. This region is found in the C-terminal section of these proteins; about 80 residues to the C-terminal of an ATP-binding site motif 'A' (P-loop) (see <PDOC00017>).

Consensus pattern[ST]-[LIVMF]-x-[LIVM]-x-D-E-[LIVMFY]-[GC]-[RKH]-G-[GST]- x(4)-G Sequences known to belong to this class detected by the pattern ALL, except for mutS2.

- [1] Modrich P. Annu. Rev. Biochem. 56:435-466(1987).
- [2] Haber L.T., Walker G.C. EMBO J. 10:2707-2715(1991).
- [3] New L., Liu K., Crouse G.F. Mol. Gen. Genet. 239:97-108(1993).
- [4] Eisen J.A. Nucleic Acids Res. 26:4291-4300(1998).

971. MutS family, N-terminal putative DNA binding domain (MutS_N)

This family consists of the N-terminal region of proteins in the mutS family of DNA mismatch repair proteins and is found associated with MutS_C located in the C-terminal region. The mutS family of proteins is named after the salmonella typhimurium MutS protein involved in mismatch repair; other members of the family included the eukaryotic MSH 1,2,3,4,5 and 6 proteins. These have various roles in DNA repair and recombination. Human MSH has been implicated in non-polyposis colorectal carcinoma (HNPCC) and is a mismatch binding protein [2]. The aligned region corresponds in part with domains A1, A2 (which may bind DNA) and B (which binds dsDNA in vitro) from *T. thermophilus* MutS as characterised in [1].

Number of members: 43

972. Domain in Myosin and Kinesin Tails (MyTH4)

0969990-101300

5

10

15

20

25

30

Consensus pattern R-G-[LIVMF]-E-x(15)-[QESM]-R-x-C-G-[LIVM]-C [The two C's are nickel ligands] Sequences known to belong to this class detected by the pattern ALL.

[1] Menon N.K., Robbins J., Peck H.D. Jr., Chatelus C.Y., Choi E.-S., Przybyla A.E. J. Bacteriol. 172:1969-1977(1990).

[2] Volbeda A., Charon M.-H., Piras C., Hatchikian E.C., Frey M., Fontecilla-Camps J.C. Nature 373:580-587(1995).

[3] Eidsness M.K., Scott R.A., Prickrill B., der Vartanian D.V., LeGall J., Moura I., Moura J.J.G., Peck H.D. Jr. Proc. Natl. Acad. Sci. U.S.A. 86:147-151(1989).

[4] Tran-Betcke A., Warnecke U., Boecker C., Zaborosch C., Friedrich B. J. Bacteriol. 172:2920-2929(1990).

This sub-family represents a carboxyl terminal extension of oxidored_q1. Only NADH-Ubiquinone chain 5 from chloroplasts are in this family. This sub-family is part of complex I which catalyses the transfer of two electrons from NADH to ubiquinone in a reaction that is associated with proton translocation across the membrane.

[1] Walker JE; Medline: 93110040 The NADH:ubiquinone oxidoreductase (complex I) of respiratory chains." Q Rev Biophys 1992;25:253-324.

This sub-family represents an amino terminal extension of oxidored_q1. Only NADH-Ubiquinone chain 5 and eubacterial chain L are in this family. This sub-family is part of

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100

546

- 5

10

- 15

- 20

25

SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 3 FAMILY.

- 30

- G.L., Koessel H.; "Complete sequence of the maize chloroplast genome: gene content,

hotspots of divergence and fine tuning of genetic information by transcript editing."; J. Mol. Biol. 251:614-628(1995).

980. PAC: PAC motif

- 5 PAC motif occurs C-terminal to a subset of all known PAS motifs. It is proposed to contribute to the PAS domain fold [3]. Number of members: 181

[1] Medline: 97446881 PAS domain S-boxes in archaea, bacteria and sensors for oxygen and redox. Zhulin IB, Taylor BL, Dixon R; Trends Biochem Sci 1997;22:331-333.

- 10 [2] Medline: 95275818. 1.4 A structure of photoactive yellow protein, a cytosolic photoreceptor: unusual fold, active site, and chromophore. Borgstahl GE, Williams DR, Getzoff ED; Biochemistry 1995;34:6278-6287.

[3] Medline: 98044337. PAS: a multifunctional domain family comes to light. Ponting CP, Aravind L; Curr Biol 1997;7:674-677.

15 981. PARP: Poly(ADP-ribose) polymerase catalytic region.

Poly(ADP-ribose) polymerase catalyses the covalent attachment of ADP-ribose units from NAD⁺ to itself and to a limited number of other DNA binding proteins, which decreases their affinity for DNA. Poly(ADP-ribose) polymerase is a regulatory component induced by DNA damage.

20 The carboxyl-terminal region is the most highly conserved region of the protein. Experiments have shown that a carboxyl 40 kDa fragment is still catalytically active [2]. Number of members: 19

25 [1] Medline: 96353841 Structure of the catalytic fragment of poly(AD-ribose) polymerase from chicken. Ruf A, Mennissier de Murcia J, de Murcia G, Schulz GE; Proc Natl Acad Sci U S A 1996;93:7481-7485.

- 30 [2] Medline: 93293867 The carboxyl-terminal domain of human poly(ADP-ribose) polymerase. Overproduction in Escherichia coli, large scale purification, and characterization. Simonin F, Hofferer L, Panzeter PL, Muller S, de Murcia G, Althaus FR; J Biol Chem 1993;268:13454-13461.

982. PC_rep: Proteasome/cyclosome repeat

[1] Medline: 97348748 A repetitive sequence in subunits of the 26S proteasome and 20S cyclosome (anaphase-promoting complex). Lupas A, Baumeister W, Hofmann K; Trends Biochem Sci 1997;22:195-196.

Number of members: 112

5

983. Peptidase_M1: Peptidase family M1

Members of this family are aminopeptidases. The members differ widely in specificity, hydrolysing acidic, basic or neutral N-terminal residues. This family includes leukotriene-A4 hydrolase Swiss:P09960, this enzyme also has an aminopeptidase activity [1]. Number of members: 72

10

[1] Medline: 95405261 Evolutionary families of metallopeptidases. Rawlings ND, Barrett AJ; Meth Enzymol 1995;248:183-228.

15

984. Neutral zinc metallopeptidases, zinc-binding region signature (Peptidase_M8) PROSITE cross-reference(s) PS00142; ZINC_PROTEASE

The majority of zinc-dependent metallopeptidases (with the notable exception of the carboxypeptidases) share a common pattern of primary structure [1,2,3] in the part of their sequence involved in the binding of zinc, and can be grouped together as a superfamily, known as the metzincins, on the basis of this sequence similarity. They can be classified into a number of distinct families [4,E1] which are listed below along with the proteases which are currently known to belong to these families.

20

Family M1

25

- Bacterial aminopeptidase N (EC 3.4.11.2) (gene pepN).
- Mammalian aminopeptidase N (EC 3.4.11.2).
- Mammalian glutamyl aminopeptidase (EC 3.4.11.7) (aminopeptidase A). It may play a role in regulating growth and differentiation of early B-lineage cells.
- Yeast aminopeptidase yscII (gene APE2).
- Yeast alanine/arginine aminopeptidase (gene AAP1).
- Yeast hypothetical protein YIL137c.
- Leukotriene A-4 hydrolase (EC 3.3.2.6). This enzyme is responsible for the hydrolysis of an epoxide moiety of LTA-4 to form LTB-4; it has been shown that it binds zinc and is capable of peptidase activity.

30

09589980-10300

Family M2

- Angiotensin-converting enzyme (EC 3.4.15.1) (dipeptidyl carboxypeptidase I) (ACE) the enzyme responsible for hydrolyzing angiotensin I to angiotensin II. There are two forms of ACE: a testis-specific isozyme and a somatic isozyme which has two active centers.

5 Family M3

- Thimet oligopeptidase (EC 3.4.24.15), a mammalian enzyme involved in the cytoplasmic degradation of small peptides.

- Neurolysin (EC 3.4.24.16) (also known as mitochondrial oligopeptidase M or microsomal endopeptidase).

10 - Mitochondrial intermediate peptidase precursor (EC 3.4.24.59) (MIP). It is involved the second stage of processing of some proteins imported in the mitochondrion.

- Yeast saccharolysin (EC 3.4.24.37) (proteinase yscD).

- Escherichia coli and related bacteria dipeptidyl carboxypeptidase (EC 3.4.15.5) (gene dcp).

15 - Escherichia coli and related bacteria oligopeptidase A (EC 3.4.24.70) (gene opdA or prlC).

- Yeast hypothetical protein YKL134c.

Family M4

- Thermostable thermolysins (EC 3.4.24.27), and related thermolabile neutral proteases (bacillolysins) (EC 3.4.24.28) from various species of Bacillus.

20 - Pseudolysin (EC 3.4.24.26) from Pseudomonas aeruginosa (gene lasB).

- Extracellular elastase from Staphylococcus epidermidis.

- Extracellular protease prt1 from Erwinia carotovora.

- Extracellular minor protease smp from Serratia marcescens.

- Vibriolysin (EC 3.4.24.25) from various species of Vibrio.

25 - Protease prtA from Listeria monocytogenes.

- Extracellular proteinase proA from Legionella pneumophila.

Family M5

- Mycolysin (EC 3.4.24.31) from Streptomyces cacaoi.

30

Family M6

- Immune inhibitor A from Bacillus thuringiensis (gene ina). Ina degrades two classes of insect antibacterial proteins, attacins and cecropins.

05689980 "101300

- Streptomyces extracellular small neutral proteases

5 - Leishmanolysin (EC 3.4.24.36) (surface glycoprotein gp63), a cell surface protease from various species of *Leishmania*.

10 - Microbial collagenase (EC 3.4.24.3) from *Clostridium perfringens* and *Vibrio*
alginolyticus.

- Serralyisin (EC 3.4.24.40), an extracellular metalloprotease from *Serratia*.

- Alkaline metalloproteinase from *Pseudomonas aeruginosa* (gene *aprA*).

15 - Secreted proteases A, B, C and G from *Erwinia chrysanthemi*.

- Yeast hypothetical protein YIL108w.

20 - Mammalian extracellular matrix metalloproteinases (known as matrixins) [5]: MMP-1 (EC 3.4.24.7) (interstitial collagenase), MMP-2 (EC 3.4.24.24) (72 Kd gelatinase), MMP-9 (EC 3.4.24.35) (92 Kd gelatinase), MMP-7 (EC 3.4.24.23) (matrylisin), MMP-8 (EC 3.4.24.34) (neutrophil collagenase), MMP-3 (EC 3.4.24.17) (stromelysin-1), MMP-10 (EC 3.4.24.22) (stromelysin-2), and MMP-11 (stromelysin-3), MMP-12 (EC 3.4.24.65) (macrophage metalloelastase).

25 - Sea urchin hatching enzyme (envelysin) (EC 3.4.24.12). A protease that allows the embryo to digest the protective envelope derived from the egg extracellular matrix.

- Soybean metalloendoproteinase 1.

30 - *Chlamydomonas reinhardtii* gamete lytic enzyme (GLE).

- Astacin (EC 3.4.24.21), a crayfish endoprotease.

- Meprin A (EC 3.4.24.18), a mammalian kidney and intestinal brush border

[illegible]

metalloendopeptidase.

- Bone morphogenetic protein 1 (BMP-1), a protein which induces cartilage and bone formation and which expresses metalloendopeptidase activity. The *Drosophila* homolog of BMP-1 is the dorsal-ventral patterning protein tolloid.
- Blastula protease 10 (BP10) from *Paracentrotus lividus* and the related protein SpAN from *Strongylocentrotus purpuratus*.
- *Caenorhabditis elegans* protein toh-2.
- *Caenorhabditis elegans* hypothetical protein F42A10.8.
- Choriolysins L and H (EC 3.4.24.67) (also known as embryonic hatching proteins LCE and HCE) from the fish *Oryzias latipes*. These proteases participate in the breakdown of the egg envelope, which is derived from the egg extracellular matrix, at the time of hatching.

Family M12B

- Snake venom metalloproteinases [6]. This subfamily mostly groups proteases that act in hemorrhage. Examples are: adamalysin II (EC 3.4.24.46), atrolysin C/D (EC 3.4.24.42), atrolysin E (EC 3.4.24.44), fibrolase (EC 3.4.24.72), trimere lysin I (EC 3.4.25.52) and II (EC 3.4.25.53).
- Mouse cell surface antigen MS2.

Family M13

- Mammalian neprilysin (EC 3.4.24.11) (neutral endopeptidase) (NEP).
- Endothelin-converting enzyme 1 (EC 3.4.24.71) (ECE-1), which process the precursor of endothelin to release the active peptide.
- Kell blood group glycoprotein, a major antigenic protein of erythrocytes. The Kell protein is very probably a zinc endopeptidase.
- Peptidase O from *Lactococcus lactis* (gene pepO).

Family M27

- Clostridial neurotoxins, including tetanus toxin (TeTx) and the various botulinum toxins (BoNT). These toxins are zinc proteases that block neurotransmitter release by proteolytic cleavage of synaptic proteins such as synaptobrevins, syntaxin and SNAP-25 [7,8].

Family M30

- Staphylococcus hyicus neutral metalloprotease.

Family M32

- 5 - Thermostable carboxypeptidase 1 (EC 3.4.17.19) (carboxypeptidase Taq), an enzyme from *Thermus aquaticus* which is most active at high temperature.

Family M34

- 10 - Lethal factor (LF) from *Bacillus anthracis*, one of the three proteins composing the anthrax toxin.

Family M35

- 15 - Deuterolysin (EC 3.4.24.39) from *Penicillium citrinum* and related proteases from various species of *Aspergillus*.

Family M36

- Extracellular elastinolytic metalloproteinases from *Aspergillus*.

20 From the tertiary structure of thermolysin, the position of the residues acting as zinc ligands and those involved in the catalytic activity are known. Two of the zinc ligands are histidines which are very close together in the sequence; C-terminal to the first histidine is a glutamic acid residue which acts as a nucleophile and promotes the attack of a water molecule on the carbonyl carbon of the substrate. A signature pattern which includes the two histidine and the glutamic acid residues is sufficient to detect this superfamily of
25 proteins.

Consensus pattern[GSTALIVN]-x(2)-H-E-[LIVMFYW]-{DEHRKP}-H-x-
[LIVMFYWGSPQ]

[The two H's are zinc ligands] [E is the active site residue]

30 Sequences known to belong to this class detected by the patternALL, except for members of families M5, M7 and M11.

Other sequence(s) detected in SWISS-PROT57; including *Neurospora crassa* conidiation-specific protein 13 which could be a zinc-protease.

[1]Jongeneel C.V., Bouvier J., Bairoch A. FEBS Lett. 242:211-214(1989).

09689980-101300

- 5

10

This family includes PHO-4 from *Neurospora crassa* which is a Na(+)-phosphate symporter [1]. This family also contains the leukemia virus receptor Swiss:Q08344. Number of members: 41

- 15

986. Photosynthetic reaction center proteins signature (photoRC)

PROSITE cross-reference(s): PS00244; REACTION_CENTER

20

30

Consensus pattern[NQH]-x(4)-P-x-H-x(2)-[SAG]-x(11)-[SAGC]-x-H-[SAG](2)

[The first H is a magnesium ligand] [The second H is a iron ligand]

Sequences known to belong to this class detected by the patternALL, except

for broad bean psbA which has Gln instead of the second His.

[1]Michel H., Deisenhofer J. Biochemistry 27:1-7(1988).

[2]Barber J. Trends Biochem. Sci. 12:321-326(1987).

5

987. phytochrome: Phytochrome region

This family contains a region specific to phytochrome proteins. Number of members:

145

10 988. PI3K_C2: C2 domain

Phosphoinositide 3-kinase region postulated to contain a C2 domain. Outlier of C2 family.

Number of members: 39

[1] Medline: 97388296 Using structure to define the function of phosphoinositide 3-kinase family members. Domin J, Waterfield MD; FEBS Lett 1997;410:91-95.

[2] Medline: 97398940 Phosphoinositide 3-kinases: a conserved family of signal transducers. Vanhaesebroeck B, Leevers SJ, Panayotou G, Waterfield MD; Trends Biochem Sci 1997;22:267-272.

20 989. PI3Ka: Phosphoinositide 3-kinase family, accessory domain (PIK domain)

PIK domain is conserved in all PI3 and PI4-kinases. Its role is unclear but it has been suggested [2] to be involved in substrate presentation.

Number of members: 47

25 [1] Medline: 97388296 Using structure to define the function of phosphoinositide 3-kinase
family members. Domin J, Waterfield MD; FEBS Lett 1997;410:91-95.

[2] Medline: 94069320 Phosphatidylinositol 4-kinase: gene structure and requirement for yeast cell viability. Flanagan CA, Schnieders EA, Emerick AW, Kunisawa R, Admon A, Thorner J; Science 1993;262:1444-1448.

30

990. P-II protein signatures

PROSITE cross-reference(s): PS00496; PII_GLNB_UMP, PS00638; PII_GLNB_CTER

Figure 1 consists of seven sub-graphs labeled (a) through (g), each showing a time course of a different physiological variable over a 10-minute period. The x-axis for all graphs represents time in minutes, from 0 to 10. The y-axis for all graphs represents the magnitude of the variable, ranging from 0 to 100. Each graph shows a baseline period (from 0 to 5 minutes) and an intervention period (from 5 to 10 minutes). The variables are: (a) Heart rate (b/min), (b) Blood pressure (mmHg), (c) Cardiac output (l/min), (d) Stroke volume (ml), (e) Stroke volume index (ml/m²), (f) Stroke volume index (ml/m²), and (g) Stroke volume index (ml/m²). The graphs show that heart rate, blood pressure, and cardiac output increase during the intervention period, while stroke volume and stroke volume index remain relatively stable.

The P-II protein (gene *glnB*) is a bacterial protein important for the control of glutamine synthetase [1,2,3]. In nitrogen-limiting conditions, when the ratio of glutamine to 2-ketoglutarate decreases, P-II is uridylylated on a tyrosine residue to form P-II-UMP. P-II-UMP allows the deadenylation of glutamine synthetase (GS), thus activating the enzyme. Conversely, in nitrogen excess, P-II-UMP is deuridylylated and then promotes the adenylation of GS. P-II also indirectly controls the transcription of the GS gene (*glnA*) by preventing NR-II (*ntrB*) to phosphorylate NR-I (*ntrC*) which is the transcriptional activator of *glnA*. Once P-II is uridylylated, these events are reversed.

P-II is a protein of about 110 amino acid residues extremely well conserved. The tyrosine which is uridylylated is located in the central part of the protein.

In cyanobacteria, P-II seems to be phosphorylated on a serine residue rather than being uridylylated.

In methanogenic archaeobacteria, the nitrogenase iron protein gene (*nifH*) is followed by two open reading frames highly similar to the eubacterial P-II protein [4]. These proteins could be involved in the regulation of nitrogen fixation.

In the red alga, *Porphyra purpurea*, there is a *glnB* homolog encoded in the chloroplast genome.

Other proteins highly similar to *glnB* are:

- *Bacillus subtilis* protein *nrgB* [5].
- *Escherichia coli* hypothetical protein *ybaI* [6].

Two signature patterns were developed for P-II protein. The first one is a conserved stretch (in eubacteria) of six residues which contains the uridylylated tyrosine, the other is derived from a conserved region in the C-terminal part of the P-II protein.

Consensus pattern Y-[KR]-G-[AS]-[AE]-Y [The second Y is uridylylated]

Sequences known to belong to this class detected by the pattern ALL *glnB*'s from eubacteria.

Consensus pattern[ST]-x(3)-G-[DY]-G-[KR]-[IV]-[FW]-[LIVM]-x(2)-[LIVM]

[1]Magasanik B. Biochimie 71:1005-1012(1989).

[2]Holtel A., Merrick M. Mol. Gen. Genet. 215:134-138(1988).

5 [3]Cheah E., Carr P.D., Suffolk P.M., Vasuvedan S.G., Dixon N.E., Ollis D.L. Structure
2:981-990(1994).

[4]Sibold L., Henriquet M., Possot O., Aubert J.-P. Res. Microbiol. 142:5-12(1991).

[5]Wray L.V. Jr., Atkinson M.R., Fisher S.H. J. Bacteriol. 176:108-114(1994).

[6]Allikmets R., Gerrard B.C., Court D., Dean M.C. Gene 136:231-236(1993).

10

991. PIP5K: Phosphatidylinositol-4-phosphate 5-Kinase

This family contains a region from the common kinase core found in the type I
phosphatidylinositol-4-phosphate 5-kinase (PIP5K) family as described in [1]. The family
consists of various type I, II and III PIP5K enzymes. PIP5K catalyses the formation of
15 phosphoinositol-4,5-bisphosphate via the phosphorylation of phosphatidylinositol-4-
phosphate a precursor in the phosphoinositide signaling pathway. Number of members: 33

[1] Medline: 98204859. Type I phosphatidylinositol-4-phosphate 5-kinases. Cloning of the
third isoform and deletion/substitution analysis of members of this novel lipid kinase family.

20 Ishihara H, Shibasaki Y, Kizuki N, Wada T, Yazaki Y, Asano T, Oka Y; J Biol Chem
1998;273:8741-8748.

[2] Medline: 97115834 Type I phosphatidylinositol-4-phosphate 5-kinases are distinct
members of this novel lipid kinase family. Loijens JC, Anderson RA; J Biol Chem 1996
20;271:32937-32943.

25

992. PolyA_pol: Poly A polymerase family

This family includes nucleic acid independent RNA polymerases, such as Poly(A)
polymerase, which adds the poly (A) tail to mRNA EC:2.7.7.19. This family also includes the
tRNA nucleotidyltransferase that adds the CCA to the 3' of the tRNA

30 EC:2.7.7.25. Number of members: 31

[1] Medline: 93066242 Identification of the gene for an Escherichia coli poly(A) polymerase.
Cao GJ, Sarkar N; Proc Natl Acad Sci U S A 1992;89:10380-10384.

OFFICE OF THE ATTORNEY GENERAL

993. Photosystem I psaA and psaB proteins signature (psaA_psaB)

PROSITE cross-reference(s)PS00419; PHOTOSYSTEM_I_PSAAB

Photosystem I (PSI) [1] is an integral membrane protein complex that uses light energy to mediate electron transfer from plastocyanin to ferredoxin. PSI is found in the chloroplast of plants and cyanobacteria. The electron transfer components of the reaction center of PSI are a primary electron donor P-700 (chlorophyll dimer) and five electron acceptors: A0 (chlorophyll), A1 (a phylloquinone) and three 4Fe-4S iron-sulfur centers: Fx, Fa, and Fb.

PsaA and psaB, two closely related proteins, are involved in the binding of P700, A0, A1, and Fx. psaA and psaB are both integral membrane proteins of 730 to 750 amino acids that seem to contain 11 transmembrane segments. The Fx 4Fe-4S iron-sulfur center is bound by four cysteines; two of these cysteines are provided by the psaA protein and the two others by psaB. The two cysteines in both proteins are proximal and located in a loop between the ninth and tenth transmembrane segments. A leucine zipper motif seems to be present [2] downstream of the cysteines and could contribute to dimerization of psaA/psaB.

The signature pattern for these proteins is based on the perfectly conserved region that includes the two iron-sulfur binding cysteines.

Consensus pattern C-D-G-P-G-R-G-G-T-C [The two C's bind the iron-sulfur center]

[1]Golbeck J.H. Biochim. Biophys. Acta 895:167-204(1987).

[2]Webber A.N., Malkin R. FEBS Lett. 264:1-14(1990).

994. PSBH: Photosystem II 10 kDa phosphoprotein

This protein is phosphorylated in a light dependent reaction.

Number of members: 20

995. PsbJ

This family consists of the photosystem II reaction center protein PsbJ from plants and Cyanobacteria. In Synechocystis sp. PCC 6803 PsbJ regulates the number of photosystem II centers in thylakoid membranes, it is a predicted 4kDa protein with one membrane spanning domain [1]. Number of members: 20

0963980-10100
DATE OF "0963980"

[1] Medline: 93131892. Genetic and immunological analyses of the cyanobacterium *Synechocystis* sp. PCC 6803 show that the protein encoded by the *psbJ* gene regulates the number of photosystem II centers in thylakoid membranes. Lind LK, Shukla VK, Nyhus KJ, Pakrasi HB; J Biol Chem 1993;268:1575-1579.

5

996. PSBT: Photosystem II reaction centre T protein

The exact function of this protein is unknown. It probably consists of a single transmembrane spanning helix. The Swiss:P37256 protein, appears to be (i) a novel photosystem II subunit and (ii) required for maintaining optimal photosystem II activity under adverse growth conditions [1]. Number of members: 17

10

[1] Medline: 94298765. The chloroplast *ycf8* open reading frame encodes a photosystem II polypeptide which maintains photosynthetic activity under adverse growth conditions. Monod C, Takahashi Y, Goldschmidt-Clermont M, Rochaix JD; EMBO J 1994;13:2747-2754.

15

997. PSI_8. PHOTOSYSTEM I REACTION CENTRE SUBUNIT VIII. Synonym(s)PSI-I. Gene name(s)PSAI. From *Hordeum vulgare* (Barley). Encoded on Chloroplast. Taxonomy Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; *Hordeum*.

20

MAY HELP IN THE ORGANIZATION OF THE PSAL SUBUNIT. BELONGS TO THE PSAI FAMILY.

[1] SEQUENCE FROM N.A. MEDLINE; 90036933. Scheller H.V., Okkels J.S., Hoej P.B., Svendsen I., Roepstorff P., Moeller B.L.; "The primary structure of a 4.0-kDa photosystem I polypeptide encoded by the chloroplast *psaI* gene."; J. Biol. Chem. 264:18402-18406(1989).

25

998. PSI_PsaJ: Photosystem I reaction centre subunit IX / PsaJ

This family consists of the photosystem I reaction centre subunit IX or PsaJ from various organisms including *Synechocystis* sp. (strain pcc 6803), *Pinus thunbergii* (green pine) and *Zea mays* (maize). PsaJ Swiss:P19443 is a small 4.4kDa, chloroplastal encoded, hydrophobic subunit of the photosystem I reaction complex its function is not yet fully understood [1].

30

PsaJ can be cross-linked to PsaF Swiss:P12356 and has a single predicted transmembrane

[1] Medline: 99238330. A large fraction of PsaF is nonfunctional in photosystem I complexes lacking the PsaJ subunit. Fischer N, Boudreau E, Hippler M, Drepper F, Haehnel W, Rochaix JD; *Biochemistry* 1999;38:5546-5552.

[2] Medline: 93252282. Genes encoding eleven subunits of photosystem I from the thermophilic cyanobacterium *Synechococcus* sp. Muhlenhoff U, Haehnel W, Witt H, Herrmann RG; Gene 1993;127:71-78.

999. PSII. Protein namePHOTOSYSTEM II P680 CHLOROPHYLL A APOPROTEIN.
 Synonym(s)CP-47 PROTEIN. Gene name(s)PSBB. From *Hordeum vulgare* (Barley),
 Encoded on Chloroplast. Taxonomy Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; *Hordeum*.

FUNCTION: THIS PROTEIN CONJUGATES WITH CHLOROPHYLL & CATALYZES THE PRIMARY LIGHT-INDUCED PHOTOCHEMICAL PROCESSES OF PHOTOSYSTEM II. SUBCELLULAR LOCATION: CHLOROPLAST THYLAKOID MEMBRANE. SIMILARITY: BELONGS TO THE PSBB / PSBC FAMILY.

[1] SEQUENCE FROM N.A. STRAIN=CV. SABARLIS; MEDLINE; 89240047. Andreeva A.V., Buryakova A.A., Reverdatto S.V., Chakhmakhcheva O.G., Efimov V.A.; "Nucleotide sequence of the 5.2 kbp barley chloroplast DNA fragment, containing psbB-psbH-petB-petD gene cluster."; Nucleic Acids Res. 17:2859-2860(1989).

[2] SEQUENCE FROM N.A. STRAIN=CV. SABARLIS; MEDLINE; 92207253. Efimov V.A., Andreeva A.V., Reverdatto S.V., Chakhmakhcheva O.G.; "Photosystem II of rye. Nucleotide sequence of the psbB, psbC, psbE, psbF, psbH genes of rye and chloroplast DNA regions adjacent to them."; Bioorg. Khim. 17:1369-1385(1991).

[3] SEQUENCE OF 411-420. Hinz U.G.; "Isolation of the photosystem II reaction center complex from barley. Characterization by circular dichroism spectroscopy and amino acid sequencing."; Carlsberg Res. Commun. 50:285-298(1985).

1000. QRPTase. Quinolinate phosphoribosyl transferase.

Quinolinate phosphoribosyl transferase (QPRTase) or nicotinate-nucleotide pyrophosphorylase EC:2.4.2.19 is involved in the de novo synthesis of NAD in both

prokaryotes and eukaryotes. It catalyses the reaction of quinolinic acid with 5-phosphoribosyl-1-pyrophosphate (PRPP) in the presence of Mg^{2+} to give rise to nicotinic acid mononucleotide (NaMN), pyrophosphate and carbon dioxide [1,2]. Number of members: 26.

[1]Medline: 97169443. A new function for a common fold: the crystal structure of quinolinic acid phosphoribosyltransferase. Eads JC, Ozturk D, Wexler TB, Grubmeyer C, Sacchettini JC; Structure 1997;5:47-58.

[2]Medline: 96139309. The sequencing expression, purification, and steady-state kinetic analysis of quinolinate phosphoribosyl transferase from Escherichia coli. Bhatia R, Calvo KC; Arch Biochem Biophys 1996;325:270-278.

1001. R3H domain

The name of the R3H domain comes from the characteristic spacing of the most conserved arginine and histidine residues. The function of the domain is predicted to be binding ssDNA. Number of members: 28

[1]Medline: 99003905 The R3H motif: a domain that binds single-stranded nucleic acids. Grishin NV; Trends Biochem Sci 1998;23:329-330.

1002. recF protein signatures (RecF)

The prokaryotic protein recF [1,2] is a single-stranded DNA-binding protein which also probably binds ATP. RecF is involved in DNA metabolism; it is required for recombinational DNA repair and for induction of the SOS response. RecF is a protein of about 350 to 370 amino acid residues; there is a conserved ATP-binding site motif 'A' (P-loop) in the N-terminal section of the protein as well as two other conserved regions, one located in the central section, and the other in the C-terminal section. Signature patterns were derived from these two regions.

Consensus pattern [LIVM]-x(4)-[LIF]-x(6)-[LIF]-[LVF]-x-[GE]-[GSTAD]-[PA]- x(2)-R-R-x-[FYW]-[LIVMF]-D Sequences known to belong to this class detected by the pattern ALL.

[1] Sandler S.J., Chackerian B., Li J.T., Clark A.J. Nucleic Acids Res. 20:839-845(1992).

1003. RibD C-terminal domain (RibD_C)

Number of members: 21

15 1004. Ribosomal protein L16 signatures (Ribosomal_L16)

Ribosomal protein L16 is one of the proteins from the large ribosomal subunit. In *Escherichia coli*, L16 is known to bind directly the 23S rRNA and to be located at the A site of the peptidyltransferase center. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups:

- Eubacterial L16.
- Algal and plant chloroplast L16.
- Cyanelle L16.
- **Plant mitochondrial L16.**

25 L16 is a protein of 133 to 185 amino-acid residues. As signature patterns, we selected two conserved regions in the central section of these proteins.

Consensus pattern [KR](2)-x-[GSAC]-[KRQVA]-[LIVM]-W-[LIVM]-[KR]-[LIVM]-[LFY]-[AP] Sequences known to belong to this class detected by the pattern ALL.

Consensus patternR-M-G-x-[GR]-K-G-x(4)-[FWKR] Sequences known to belong to this class detected by the patternALL.

- [1] Otaka E., Hashimoto T., Mizuta K., Suzuki K. Protein Seq. Data Anal. 5:301-313(1993).

1005. Ribosomal protein L32e signature (Ribosomal_L32E)

A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Mammalian L32 [1].
- Drosophila RP49 [2].
- Trichoderma harzianum L32 [3].
- Yeast L32e (YBL092w).
- Archaeobacterial L32e [4].

These proteins have 135 to 240 amino-acid residues. As a signature pattern, a stretch of about 20 residues located in the N-terminal part of these proteins was selected.

Consensus pattern F-x-R-x(4)-[KR]-x(2)-[KR]-[LIVMF]-x(3,5)-W-R-[KR]-x(2)-G Sequences known to belong to this class detected by the pattern ALL.

[1] Jacks C.M., Powaser C.B., Hackett P.B. Gene 74:565-570(1988).

[2] Aguade M. Mol. Biol. Evol. 5:433-441(1988).

[3] Lora J.M., Garcia I., Benitez T., Llobell A., Pintor-Toro J.A. Nucleic Acids Res. 21:3319-3319(1993).

[4] Arndt E., Scholzen T., Kroemer W., Hatakeyama T., Kimura M. Biochimie 73:657-668(1991).

1006. (Ribosomal_S3) Ribosomal protein S3 signature

PROSITE: PDOC00474. PROSITE cross-reference(s) PS00548; RIBOSOMAL_S3

Ribosomal protein S3 is one of the proteins from the small ribosomal subunit. In Escherichia coli, S3 is known to be involved in the binding of initiator Met-tRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups:

- Eubacterial S3.
- Algal and plant chloroplast S3.
- Cyanelle S3.
- Archaeobacterial S3.
- Plant mitochondrial S3.

-Vertebrate S3.

-Insect S3.

-Caenorhabditis elegans S3 (C23G10.3).

-Yeast S3 (Rp13).

- 5 S3 is a protein of 209 to 559 amino-acid residues. A conserved region located in the C-terminal section was selected as a signature pattern.

Consensus pattern[GSTA]-[KR]-x(6)-G-x-[LIVMT]-x(2)-[NQSCH]-x(1,3)-[LIVFCA]-x(3)-
[LIV]-[DENQ]-x(7)-[LMT]-x(2)-G-x(2)-[GS]. Sequences known to belong to this class
10 detected by the patternALL, except for some mitochondrial S3.

[1]Otake E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).

1007. RimM - RimM

15 The RimM protein is essential for efficient processing of 16S rRNA [1]. The RimM protein was shown to have affinity for free ribosomal 30S subunits but not for 30S subunits in the 70S ribosomes [1]. Number of members: 14.

[1]Medline: 98083058. RimM and RbfA are essential for efficient processing of 16S rRNA in
20 Escherichia coli. Bylund GO, Wipemo LC, Lundberg LA, Wikstrom PM; J Bacteriol 1998;180:73-82.

1008. RNA_pol_A - RNA polymerase alpha subunit

25 -!- RNA polymerases catalyse the DNA dependent polymerisation of RNA. Prokaryotes contain a single RNA polymerase compared to three in eukaryotes (not including mitochondrial and chloroplast polymerases).

-!- Members of this family include: A subunit from eukaryotes, gamma subunit from cyanobacteria, beta' subunit from eubacteria, A' subunit from archaebacteria, B" from chloroplasts. Number of members: 139.

30 [1]Medline: 97066998. Structural modules of the large subunits of RNA polymerase.

Introducing archaebacterial and chloroplast split sites in the beta and beta' subunits of Escherichia coli RNA polymerase. Severinov K, Mustaev A, Kukarin A, Muzzin O, Bass I, Darst SA, Goldfarb A; J Biol Chem 1996;271:27969-27974.

09689980-101300
"09689980"

1009. RuBisCO_large - Ribulose biphosphate carboxylase large chain active site

PROSITE: PDOC00142; PROSITE cross-reference(s) PS00157; RUBISCO_LARGE

Ribulose biphosphate carboxylase (EC 4.1.1.39) (RuBisCO) [1,2] catalyzes the initial step in Calvin's reductive pentose phosphate cycle in plants as well as purple and green bacteria. It consists of a large catalytic unit and a small subunit of undetermined function. In plants, the large subunit is coded by the chloroplastic genome while the small subunit is encoded in the nuclear genome. Molecular activation of RuBisCO by CO₂ involves the formation of a carbamate with the epsilon-amino group of a conserved lysine residue. This carbamate is stabilized by a magnesium ion. One of the ligands of the magnesium ion is an aspartic acid residue close to the active site lysine [3]. A pattern was developed which includes both the active site residue and the metal ligand, and which is specific to RuBisCO large chains.

Consensus pattern G-x-[DN]-F-x-K-x-D-E [K is the active site residue] [The second D is a magnesium ligand]. Sequences known to belong to this class detected by the pattern ALL, except for *Cheilopleuria bicuspidis* RuBisCO.

[1] Miziorko H.M., Lorimer G.H. Annu. Rev. Biochem. 52:507-535(1983).

[2] Akazawa T., Takabe T., Kobayashi H. Trends Biochem. Sci. 9:380-383(1984).

[3] Andersson I., Knight S., Schneider G., Lindqvist Y., Lundqvist T., Branden C.-I., Lorimer G.H. Nature 337:229-234(1989).

1010. Rve - Integrase core domain

Integrase mediates integration of a DNA copy of the viral genome into the host chromosome. Integrase is composed of three domains. The amino-terminal domain is a zinc binding domain Integrase_Zn. This domain is the central catalytic domain. The carboxyl terminal domain that is a non-specific DNA binding domain integrase. The catalytic domain acts as an endonuclease when two nucleotides are removed from the 3' ends of the blunt-ended viral DNA made by reverse transcription. This domain also catalyses the DNA strand transfer reaction of the 3' ends of the viral DNA to the 5' ends of the integration site [1]. Number of members: 694.

[1]Medline: 95099322. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Davies DR; Science 1994;266:1981-1986.

- 5 1011. (SBP_bac_3) Bacterial extracellular solute-binding proteins, family 3 signature PROSITE: PDOC00798. PROSITE cross-reference(s) PS01039; SBP_BACTERIAL_3

Bacterial high affinity transport systems are involved in active transport of solutes across the cytoplasmic membrane. The protein components of these traffic systems include one or two transmembrane protein components, one or two membrane-associated ATP-binding proteins (ABC transporters; see <PDOC00185>) and a high affinity periplasmic solute-binding protein. The later are thought to bind the substrate in the vicinity of the inner membrane, and to transfer it to a complex of inner membrane proteins for concentration into the cytoplasm.

In gram-positive bacteria which are surrounded by a single membrane and have therefore no periplasmic region the equivalent proteins are bound to the membrane via an N-terminal lipid anchor. These homolog proteins do not play an integral role in the transport process per se, but probably serve as receptors to trigger or initiate translocation of the solute through the membrane by binding to external sites of the integral membrane proteins of the efflux system.

In addition at least some solute-binding proteins function in the initiation of sensory transduction pathways.

On the basis of sequence similarities, the vast majority of these solute-binding proteins can be grouped [1] into eight families of clusters, which generally correlate with the nature of the solute bound.

Family 3 groups together specific amino acids and opine-binding periplasmic proteins and a periplasmic homolog with catalytic activity:

-Histidine-binding protein (gene hisJ) of Escherichia coli and related bacteria. An homologous lipoprotein exists in Neisseria gonorrhoeae.

-Lysine/arginine/ornithine-binding proteins (LAO) (gene argT) of Escherichia coli and related bacteria are involved in the same transport system than hisJ. Both solute-binding proteins interact with a common membrane-bound receptor hisP of the binding protein dependent transport system HisQMP.

-Glutamine-binding proteins (gene glnH) of Escherichia coli and Bacillus stearothermophilus.

- 5

10

15

20

25

30

1013. SecA_protein. SecA protein, amino terminal region

0963801-1030

5

10

15

20

25

30

PROSITE: PDOC00429. PROSITE cross-reference(s)PS00456; NA_SOLUT_SYMP_1
PS00457; NA_SOLUT_SYMP_2 PS50283; NA SOLUTE SYMP 3

It has been shown [1,2] that integral membrane proteins that mediate the intake of a wide variety of molecules with the concomitant uptake of sodium ions (sodium symporters) can be grouped, on the basis of sequence and functional similarities into a number of distinct families. One of these families is known as the sodium:solute symporter family (SSF) and

currently consists of the following proteins:

- Mammalian Na⁺/glucose co-transporter.
- Mammalian Na⁺/myo-inositol co-transporter.
- Mammalian Na⁺/nucleoside co-transporter.
- Mammalian Na⁺/neutral amino acid co-transporter.
- Escherichia coli Na⁺/proline symporter (gene putP).
- Escherichia coli Na⁺/pantothenate symporter (gene panF).
- Escherichia coli hypothetical protein yidK.
- Escherichia coli hypothetical protein yjcG.
- Bacillus subtilis hypothetical protein ywca (ipa-31R).

These integral membrane proteins are predicted to comprise at least ten membrane spanning domains. Two conserved regions were selected as signature patterns; the first one is located in the fourth transmembrane region and the second one in a loop between two transmembrane regions in the C-terminal part of these proteins.

Consensus pattern[GS]-x(2)-[LIY]-x(3)-[LIVMFYWSTAG](10)-[LIY]-[TAV]-x(2)-G-G-[LMF]-x-[SAP]. Sequences known to belong to this class detected by the patternALL.
Consensus pattern[GAST]-[LIVM]-x(3)-[KR]-x(4)-G-A-x(2)-[GAS]-[LIVMGS]-[LIVMW]-[LIVMGAT]-G-x-[LIVMGA] Sequences known to belong to this class detected by the patternALL, except for E.coli yidK.

Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so.

[1]Reizer J., Reizer A., Saier M.H. Jr. Res. Microbiol. 141:1069-1072(1991).

[2]Reizer J., Reizer A., Saier M.H. Jr. Biochim. Biophys. Acta 1197:133-136(1994).

1017. SurE - Survival protein SurE

E. coli cells with the surE gene disrupted are found to survive poorly in stationary phase [1].

It is suggested that SurE may be involved in stress response. Yeast also contains a member of

00668800 "1017" 00668800

the family Swiss:P38254. Swiss:P30887 can complement a mutation in acid phosphatase, suggesting that members of this family could be phosphatases. Number of members: 17.

[1]Medline: 95014035. A new gene involved in stationary-phase survival located at 59 minutes on the Escherichia coli chromosome. Li C, Ichikawa JK, Ravetto JJ, Kuo HC, Fu JC, Clarke S; J Bacteriol 1994;176:6015-6022.

[2]Medline: 93046805. Complementation of Saccharomyces cerevisiae acid phosphatase mutation by a genomic sequence from the yeast Yarrowia lipolytica identifies a new phosphatase. Treton BY, Le Dall MT, Gaillardin CM; Curr Genet 1992;22:345-355.

1018. Synuclein - Synuclein

There are three types of synucleins in humans, these are called alpha, beta and gamma. Alpha synuclein has been found mutated in families with autosomal dominant Parkinson's disease. A peptide of alpha synuclein has also been found in amyloid plaques in Alzheimer's patients. Number of members: 12.

[1]Medline: 98424410. The synuclein family. Lavedan C; Genome Res 1998;8:871-880.

1019. (T-box) T-box domain signatures

PROSITE: PDOC00972. PROSITE cross-reference(s) PS01283; TBOX_1 PS01264; TBOX_2

A number of eukaryotic DNA-binding proteins contain a domain of about 170 to 190 amino acids known as the T-box domain [1,2,3] and which probably binds DNA. The T-box has first been found in the mice T locus (Brachyury) protein, a transcription factor involved in mesoderm differentiation. It has since been found in the following proteins:

- Vertebrate and invertebrate homologs of the T protein.
- Mammalian proteins TBX1 to TBX6.
- Mammalian protein TBR1 which is expressed specifically in brain.
- Xenopus laevis eomesodermin (eomes).
- Xenopus laevis Vegt (or Antipodean), a transcription factor that activates the expression of wnt-8, eomes and Brachyury.
- Chicken TbxT.
- Drosophila protein optomotor-blind (omb).
- Drosophila protein brachyenteron (byn) (also known as Trg), which is

-Drosophila protein H15.

-Caenorhabditis elegans hypothetical proteins F21H11.3, F40H6.4, T07C4.2, T07C4.6 and

ZK177.10.

Consensus patternL-W-x(2)-[FC]-x(3,4)-[NT]-E-M-[LIV](2)-T-x(2)-G-[RG]-[KRQ]

Sequences known to belong to this class detected by the patternALL, except for C.elegans ZK177.10.

Consensus pattern[LIVMYW]-H-[PADH]-[DEN]-[GS]-x(3)-G-x(2)-W-M-x(3)-[IVA]-x- F
Sequences known to belong to this class detected by the patternALL, except for *C.elegans*
tbx-12, ZK177.10 and *Drosophila* H15.

[1]Bollag R.J., Siegfried Z., Cebra-Thomas J.A., Garvey N., Davison E.M., Silver L.M. Nat. Genet. 7:383-389(1994).

[2] Agulnik S.I., Garvey N., Hancock S., Ruvinsky I., Chapman D.L., Agulnik I., Bollag R.J., Papaioannou V.E., Silver L.M. Genetics 144:249-254(1996).

[3] Papaioannou V.E. Trends Genet. 13:212-213(1997).

1020. Toprim - Toprim domain

This is a conserved region from DNA primase. This corresponds to the Toprim domain common to DnaG primases, topoisomerases, OLD family nucleases and RecR proteins [1].

Both DnaG motifs IV and V are present in the alignment, the DxD (V) motif may be involved in Mg²⁺ binding and mutations to the conserved glutamate (IV) completely abolish DnaG type primase activity [1]. DNA primase EC:2.7.7.6 is a nucleotidyltransferase it synthesizes the oligoribonucleotide primers required for DNA replication on the lagging strand of the replication fork; it can also prime the leading strand and has been implicated in cell division [2]. Number of members: 133.

[1]Medline: 98391745. Toprim--a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. Aravind L, Leipe DD, Koonin EV; *Nucleic Acids Res* 1998;26:4205-4213.

The diagram illustrates the experimental setup for studying the effect of a magnetic field on polymer growth. It shows a cross-section of a polymer film of thickness h on a substrate. A magnetic field H is applied perpendicular to the film plane. A coordinate system (x, y, z) is shown with the z -axis normal to the film. The film is divided into regions of different thicknesses: $h/2$ and $h/4$. The regions are labeled $h/2$ and $h/4$.

[2]Medline: 97368180. Cloning and analysis of the dnaG gene encoding *Pseudomonas putida* DNA primase. Szafranski P, Smith CL, Cantor CR; Biochim Biophys Acta 1997;1352:243-248.

[3]Medline: 94124015. The *Haemophilus influenzae* dnaG sequence and conserved bacterial primase motifs. Versalovic J, Lupski JR; Gene 1993;136:281-286.

1021. TraB - TraB family

pAD1 is a hemolysin/bacteriocin plasmid originally identified in *Enterococcus faecalis* DS16. It encodes a mating response to a peptide sex pheromone, cAD1, secreted by recipient bacteria. Once the plasmid pAD1 is acquired, production of the pheromone ceases--a trait related in part to a determinant designated traB. However a related protein is found in *C. elegans* Swiss:Q94217, suggesting that members of the TraB family have some more general function. Number of members: 12.

[1]Medline: 94302142. Characterization of the determinant (traB) encoding sex pheromone shutdown by the hemolysin/bacteriocin plasmid pAD1 in *Enterococcus faecalis*. An FY, Clewell DB; Plasmid 1994;31:215-221.

1022. (Transpo_mutator) Transposases, Mutator family, signature PROSITE: PDOC00770. PROSITE cross-reference(s) PS01007; TRANSPOSASE_MUTATOR

Autonomous mobile genetic elements such as transposon or insertion sequences (IS) encode an enzyme, called transposase, required for excising and inserting the mobile element. On the basis of sequence similarities, transposases can be grouped into various families. One of these families has been shown [1,2,3,E1] to consist of transposases from the following elements:

- Mutator from Maize.
- Is1201 from *Lactobacillus helveticus*.
- Is905 from *Lactococcus lactis*.
- Is1081 from *Mycobacterium bovis*.
- Is6120 from *Mycobacterium smegmatis*.
- Is406 from *Pseudomonas cepacia*.
- IsRm3 from *Rhizobium meliloti*.
- IsRm5 from *Rhizobium meliloti*.

-Is256 from *Staphylococcus aureus*.

-IsT2 from *Thiobacillus ferrooxidans*.

The maize Mutator transposase (MudrA) is a protein of 823 amino acids; the bacterial transposases listed above are proteins of 300 to 420 amino acids. These proteins contain a conserved domain of about 130 residues; a signature pattern was derived from the most conserved part of this domain.

Consensus pattern D-x(3)-G-[LIVMF]-x(6)-[STAV]-[LIVMFYW]-[PT]-x-[STAV]-x(2)-[QR]-x-C-x(2)-H. Sequences known to belong to this class detected by the pattern ALL.

[1] Eisen J.A., Benito M.-I., Walbot V. *Nucleic Acids Res.* 22:2634-2636(1994).

[2] Guilhot C., Gicquel B., Davies J., Martin C. *Mol. Microbiol.* 6:107-113(1992).

[3] Wood M.S., Byrne A., Lessie T.G. *Gene* 105:101-105(1991).

1023. Transposase_8 - Transposase

Transposase proteins are necessary for efficient DNA transposition. This family consists of various *E. coli* insertion elements and other bacterial transposases some of which are members of the IS3 family. Number of members: 58.

[1] Medline: 97324595. Genetic organization and transposition properties of IS511. D. A. Mullin, D. L. Zies, A. H. Mullin, N. Caballera & B. Ely; *Mol Gen Genet* 1997;254:456-463.

[2] Medline: 97128810. The use of an improved transposon mutagenesis system for DNA sequencing leads to the characterization of a new insertion sequence of *Streptomyces lividans* 66. J. Fischer, H. Maier, P. Viell & J. Altenbuchner; *Gene* 1996;180:81-89.

[3] Medline: 97074647. Identification and nucleotide sequence of *Rhizobium meliloti* insertion sequence ISRM6, a small transposable element that belongs to the IS3 family. S. Zekri & N. Toro; *Gene* 1996;175:43-48.

1024. tRNA_int_endo - tRNA intron endonuclease

Members of this family cleave pre tRNA at the 5' and 3' splice sites to release the intron
EC:3.1.27.9. Number of members: 8.

[1]Medline: 97344075. Properties of *H. volcanii* tRNA intron endonuclease reveal a relationship between the archaeal and eucaryal tRNA intron processing systems. Kleman-Leyer K, Armbruster DW, Daniels CJ; Cell 1997;89:839-847.

5 1025. Urease - Urease signatures

PROSITE: PDOC00133PROSITE cross-reference(s) PS01120; UREASE_1 PS00145;
UREASE_2

Urease (EC 3.5.1.5) is a nickel-binding enzyme that catalyzes the hydrolysis of urea to carbon dioxide and ammonia [1]. Historically, it was the first enzyme to be crystallized (in 1926). It is mainly found in plant seeds, microorganisms and invertebrates. In plants, urease is a hexamer of identical chains. In bacteria [2], it consists of either two or three different subunits (alpha, beta and gamma).

Urease binds two nickel ions per subunit; four histidine, an aspartate and a carbamated-lysine serve as ligands to these metals; an additional histidine is involved in the catalytic mechanism [3].

As signatures for this enzyme, a region that contains two histidine that bind one of the nickel ions and the region of the active site histidine was selected.

Consensus pattern T-[AY]-[GA]-[GAT]-[LIVM]-D-x-H-[LIVM]-H-x(3)-P [The two H's bind nickel]. Sequences known to belong to this class detected by the patternALL.

Consensus pattern[LIVM](2)-[CT]-H-[HN]-L-x(3)-[LIVM]-x(2)-D-[LIVM]-x-F-A [H is the active site residue]. Sequences known to belong to this class detected by the patternALL.

[1]Takishima K., Suga T., Mamiya G. Eur. J. Biochem. 175:151-165(1988).

[2]Mobley H.L.T., Husinger R.P. Microbiol. Rev. 53:85-108(1989).

[3]Jabri E., Carr M.B., Hausinger R.P., Karplus P.A. Science 268:998-1004(1995).

1026. Urease_beta - Urease beta subunit.

This subunit is known as alpha in *Helicobacter*. Number of members: 35.

[1]Medline: 95273988. The crystal structure of urease from *Klebsiella aerogenes*. Jabri E, Carr MB, Hausinger RP, Karplus PA; Science 1995;268:998-1004.

1027. UvrD-helicase - UvrD/REP helicase

[illegible]

The Rep family helicases are composed of four structural domains. The Rep family function as dimers. REP helicases catalyse ATP dependent unwinding of double stranded DNA to single stranded DNA. Swiss:P23478, Swiss:P08394 have large insertions near to the carboxy-terminus relative to other members of the family. Number of members: 52.

5

[1] Medline: 97433075. Major domain swiveling revealed by the crystal structures of complexes of E. coli Rep helicase bound to single-stranded DNA and ADP. Korolev S, Hsieh J, Gauss GH, Lohman TM, Waksman G; Cell 1997;90:635-647.

10 1028. V-type ATPase 116kDa subunit family (V_ATPase_sub_a)

This family consists of the 116kDa V-type ATPase (vacuolar (H⁺)-ATPases) subunits, as well as V-type ATP synthase subunit i. The V-type ATPases family are proton pumps that acidify intracellular compartments in eukaryotic cells for example yeast central vacuoles, clathrin-coated and synaptic vesicles. They have important roles in membrane trafficking processes [1]. The 116kDa subunit (subunit a) in the V-type ATPase is part of the V0 functional domain responsible for proton transport. The a subunit is a transmembrane glycoprotein with multiple putative transmembrane helices. It has a hydrophilic amino terminal and a hydrophobic carboxy terminal [1,2]. It has roles in proton transport and assembly of the V-type ATPase complex [1,2]. This subunit is encoded by two homologous gene in yeast VPH1 and STV1 [2].

Number of members: 27

[1] Forgac M; Medline: 99240666 "Structure and properties of the vacuolar (H⁺)-ATPases." J Biol Chem 1999;274:12951-12954.

[2] Forgac M; Medline: 99270697 "Structure and properties of the clathrin-coated vesicle and yeast vacuolar V-ATPases." J Bioenerg Biomembr 1999;31:57-65.

1029. Viral (Superfamily 1) RNA helicase (Viral_helicase1)

Number of members: 260

[1] Koonin EV, Dolja VV; Medline: 94094568 "Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences." Crit Rev Biochem Mol Biol 1993;28:375-430.

1030. Vesicular monoamine transporter (VMAT)

This family consists of various vesicular amine transporters with 12 transmembrane helices.

These included vesicular acetylcholine transporters (VACHT) [3], and vesicular monoamine transporters (VMATs) [1,2] isoforms 1 adrenal and 2 brain (VMAT1 and VMAT2).

These proteins transport biogenic amines into synaptic vesicles or chromaffin granules [4].

VMATs pack monoamine neurotransmitters into secretory vesicles for regulated exocytotic release, they also protect against the parkinsonian neurotoxins MPP+ by transporting it into vesicles preventing it from acting on mitochondria [1].

Also in the family is *C. elegans* UNC-17 a putative vesicular acetylcholine transporter mutations in UNC-17 cause impaired neuromuscular function, giving rise to jerky or uncoordinated movement, [4].

Number of members: 15

[1] Krantz DE, Peter D, Liu Y, Edwards RH; Medline: 97197857 Phosphorylation of a vesicular monoamine transporter by casein kinase II." J Biol Chem 1997;272:6752-6759.

[2] Erickson JD, Varoqui H, Schafer MK, Modi W, Diebler MF, Weihe E, Rand J, Eiden LE, Bonner TI, Usdin TB; Medline: 94350930 Functional identification of a vesicular acetylcholine transporter and its expression from a 'cholinergic' gene locus." J Biol Chem 1994;269:21929-21932.

[3] Erickson JD, Schafer MK, Bonner TI, Eiden LE, Weihe E; Medline: 96209876 Distinct pharmacological properties and distribution in neurons and endocrine cells of two isoforms of the human vesicular monoamine transporter." Proc Natl Acad Sci U S A 1996;93:5166-5171.

[4] Alfonso A, Grundahl K, Duerr JS, Han HP, Rand JB; Medline: 3342494 The *Caenorhabditis elegans* unc-17 gene: a putative vesicular acetylcholine transporter." Science 1993;261:617-619.

1031. WW/rsp5/WWP domain signature and profile. Cross-reference(s): PS01159; WW_DOMAIN_1; PS50020; WW_DOMAIN_2

09689980 "101300

The WW domain [1-4,E1] (also known as rsp5 or WWP) has been originally discovered as a short conserved region in a number of unrelated proteins, among them dystrophin, the gene responsible for Duchenne muscular dystrophy. The domain, which spans about 35 residues, is repeated up to 4 times in some proteins. It has been shown [5] to bind proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. It appears to contain beta-strands grouped around four conserved aromatic positions; generally Trp. The name WW or WWP derives from the presence of these Trp as well as that of a conserved Pro. It is frequently associated with other domains typical for proteins in signal transduction processes.

Proteins containing the WW domain are listed below.

--Dystrophin, a multidomain cytoskeletal protein. Its longest alternatively spliced form consists of an N-terminal actin-binding domain, followed by 24 spectrin-like repeats, a cysteine-rich calcium-binding domain and a C-terminal globular domain. Dystrophin form tetramers and is thought to have multiple functions including involvement in membrane stability, transduction of contractile forces to the extracellular environment and organization of membrane specialization. Mutations in the dystrophin gene lead to muscular dystrophy of Duchenne or Becker type. Dystrophin contains one WW domain C-terminal of the spectrin-repeats.

--Utrophin, a dystrophin-like protein of unknown function.

--Vertebrate YAP protein is a substrate of an unknown serine kinase. It binds to the SH3 domain of the Yes oncoprotein via a proline-rich region. This protein appears in alternatively spliced isoforms, containing either one or two WW domains [6].

--Mouse NEDD-4 plays a role in the embryonic development and differentiation of the central nervous system. It contains 3 WW modules followed by a HECT domain. The human ortholog contains 4 WW domains, but the third WW domain is probably spliced resulting in an alternate NEDD-4 protein with only 3 WW modules [3].

--Yeast RSP5 is similar to NEDD-4 in its molecular organization. It contains an N-terminal C2 domain (see <PDOC00380>), followed by a histidine-rich region, 3 WW domains and a HECT domain.

--Rat FE65, a transcription-factor activator expressed preferentially in liver. The activator domain is located within the N-terminal 232 residues of FE65, which also contain the WW domain.

--Yeast ESS1/PTF1, a putative peptidyl prolyl cis-trans isomerase from family ppiC (see <PDOC00840>). A related protein, dodo (gene dod) exists in *Drosophila* and in mammals (gene PIN1).

--Tobacco DB10 protein. The WW domain is located N-terminal to the region with similarity to ATP-dependent RNA helicases.

--IQGAP, a human GTPase activating protein acting on ras. It contains an N-terminal domain similar to fly muscle mp20 protein and a C-terminal ras GTPase activator domain.

--Yeast pre-mRNA processing protein PRP40, *Caenorhabditis elegans* ZK1098.1 and fission yeast SpAC13C5.02 are related proteins with similarity to MYO2-type myosin, each containing two WW-domains at the N-terminus.

--*Caenorhabditis elegans* hypothetical protein C38D4.5, which contains one WW module, a PH domain (see <PDOC50003>) and a C-terminal phosphatidylinositol 3-kinase domain.

--Yeast hypothetical protein YFL010c.

For the sensitive detection of WW domains, a profile was developed which spans the whole homology region as well as a pattern.

Description of pattern(s) and/or profile(s):

Consensus pattern W-x(9,11)-[VFY]-[FYW]-x(6,7)-[GSTNE]-[GSTQCR]-[FYW]-x(2)-P.

[1] Bork P., Sudol M. Trends Biochem. Sci. 19:531-533(1994).

[2] Andre B., Springael J.Y. Biochem. Biophys. Res. Commun. 205:1201-1205(1994).

[3] Hofmann K.O., Bucher P. FEBS Lett. 358:153-157(1995).

[4] Sudol M., Chen H.I., Bougeret C., Einbond A., Bork P. FEBS Lett. 369:67-71(1995).

[5] Chen H.I., Sudol M. Proc. Natl. Acad. Sci. U.S.A. 92:7819-7823(1995).

[6] Sudol M., Bork P., Einbond A., Kastury K., Druck T., Negrini M., Huebner K., Lehman D. J. Biol. Chem. 270:14733-14741(1995).

1032. XPA protein signatures. cross-reference(s): XPA_1 PROSITE PS00752; PS00753; XPA_2.

Xeroderma pigmentosum (XP) [1] is a human autosomal recessive disease, characterized by a high incidence of sunlight-induced skin cancer. People's skin cells with this condition are hypersensitive to ultraviolet light, due

0968980 "101300"

to defects in the incision step of DNA excision repair. There are a minimum of seven genetic complementation groups involved in this pathway: XP-A to XP-G. XP-A is the most severe form of the disease and is due to defects in a 30 Kd nuclear protein called XPA (or XPAC) [2].

5

The sequence of the XPA protein is conserved from higher eukaryotes [3] to yeast (gene RAD14) [4]. XPA is a hydrophilic protein of 247 to 296 amino-acid residues which has a C4-type zinc finger motif in its central section.

10

Two signature were developed patterns for XPA proteins. The first corresponds to the zinc finger region, the second to a highly conserved region located some 12 residues after the zinc finger region.

Consensus pattern C-x-[DE]-C-x(3)-[LIVMF]-x(1,2)-D-x(2)-L-x(3)-F-x(4)-C-x(2)-C

15

Consensus pattern [LIVM](2)-T-[KR]-T-E-x-K-x-[DE]-Y-[LIVMF](2)-x-D-x-[DE]

[1] Tanaka K., Wood R.D. Trends Biochem. Sci. 19:83-86(1994).

[2] Miura N., Miyamoto I., Asahina H., Satokata I., Tanaka K., Okada Y. J. Biol. Chem. 266:19786-19789(1991).

20

[3] Shimamoto T., Kohno K., Tanaka K., Okada Y. Biochem. Biophys. Res. Commun. 181:1231-1237(1991).

[4] Bankmann M., Prakash L., Prakash S. Nature 355:555-558(1992).

1033. YCF9

25

This family consists of the hypothetical protein product of the YCF9 gene from chloroplasts and cyanobacteria. Number of members: 16

1034. (DUF15)

30

It is highly conserved between eubacteria and eukaryotes.

Number of members: 30

1035. Lumenal portion of Cytochrome b559, alpha (gene psbE) subunit. (cytochr_b559a)

DOCTOT"08662360

This family is the lumenal portion of cytochrome b559 alpha chain, matches to this family should be accompanied by a match to the cytochr_b559 family also. The Prosite pattern pattern matches the transmembrane region of the cytochrome b559 alpha and beta subunits.

5 Number of members: 16

A. Asparaginase 2

10 Asparaginase II (L-asparagine aminohydrolase II) is an extracellular protein that may be associated with the cell wall and whose expression is affected by the availability of nitrogen. Asparaginase II catalyzes the reaction of L-Asparagine + H₂O = L-Aspartate + NH₃. As many leukemias have high requirements for aspartic acid, asparaginase II proteins are useful
15 as reagents for screening compounds for activity as leukemia chemotherapy products. Asparaginase II protein can also be over- or under-expressed to alter amino acid content in plant tissues or to modify nitrogen fixation and/or nitrogen metabolism in plants.

Ref: Bon et al. (1997) Appl Biochem Biotechnol 63-65: 203-12

B. Chloroa b-bind

20 Chlorophyll a-b binding proteins are located in the thylakoid membranes of the chloroplast and bind chlorophyll a and chlorophyll b, thereby triggering a chemical reaction
25 (photosynthesis). These proteins are useful in controlling the rate, efficiency and/or output of photosynthesis. Overexpression of chlorophyll a-b binding proteins is expected to increase the rate of photosynthesis.

Ref: Leutwiler et al. (1986) Nucleic Acids Res 14: 4051-64

30 Brandt et al. (1992) Plant Mol Biol 19: 699-703

C. DMRL synthase

DMRL Synthase (6,7-Dimethyl-8-Ribityllumazine Synthase) catalyzes the last step in riboflavin (Vitamin B₂) synthesis, condensing 5-amino-6-(1'-D)-ribityl-amino-2,4(1H, 3H)-Pyrimidinedione with L-3,4-Dihydroxy-2-Butanone 4-Phosphate producing 6,7-Dimethyl-8-(1-D-Ribityl)Luminazine . The enzyme forms a homopentamer. Engineering of these proteins or those with homologous sequences/structures may allow control of the amounts of vitamin B₂ available in plants and/or accumulation of pigment, as well as altering reactions requiring hydrogen ion carriers/transmitters.

Ref: Garcia-Ramirez et al. (1995) J Biol Chem **270**: 23801-7

D. E1_N

These proteins are ATP-dependent DNA helicases that are required for initiation of viral DNA replication. They form a complex with the viral E2 protein. The E1-E2 complex binds to the replication origin that contains binding sites for both proteins. The majority of sequences known for this group of proteins are from various papillomaviruses, a type of double stranded DNA virus. In plants, the prototype double stranded DNA virus is Cauliflower Mosaic virus (CaMV). Manipulation of these proteins, especially to produce variant proteins that form non-productive complexes, enables production of plants that are resistant to infection by double stranded DNA viruses.

Ref: Yang et al. (1993) PNAS USA **90**: 5086-90

Ustav and Stenlund (1991) EMBO J **10**: 449-57

Callaway et al. (1996) Mol Plant Microbe Interact **9**: 810-8

E. EF1_G

Elongation Factor-1 is composed of four subunits: alpha, beta, delta and gamma. Gamma subunits are presumed to play a role in anchoring the complex to other cellular components. Studies of EF-1 genes in plants suggests that different forms of the EF-1 subunits may be expressed in particular organs or in response to stress. Manipulation of the activity of these proteins, either by altered expression level or by structural mutation, may result in the accumulation of a particular protein in a chosen organ or allow production of particular proteins during stress conditions.

Ref: Kinzy et al. (1994) NAR 22: 2703-7
Dunn et al. (1993) Plant Mol Biol 23: 221-5
Aguilar et al. (1991) Plant Mol Biol 17: 351-60

5

F. ENV_polyprotein

This family comprises the envelope or coat proteins known from a number of different retroviruses. In mammalian species, retroviruses are responsible for diseases such as leukemia and HIV. In plants, retroviruses are known in both monocot (e.g. Zeon-1) and dicot (e.g. Arabidopsis and tobacco) species and have been shown to induce mutant alleles at new loci. Engineering of plant ENV proteins may allow mobilization or targeting of endogenous or introduced retroviruses, in essence generating a new method for mutant production, gene tagging and the like.

Ref: Mamoun et al (1990) J Virol 64: 4180-8
Grandbastien et al. (1989) Nature 337: 376-80
Wright and Voytas (1998) Genetics 149: 703-15

G. Glycosyl_hydr9

Proteins having this domain (previously known as the glycosyl hydrolase family 5 domain) catalyze the endohydrolysis of 1,4- β -D-glucosidic linkages in cellulose. Numerous plant proteins with this domain exist and are expressed in an organ specific manner. They are involved in the fruit ripening process, in cell elongation and plant reproduction. Modulation of the activity of these proteins, either by over- or under-expression or by mutation of the polypeptide, could be used to affect post-harvest physiology (e.g. rate of ripening) or for engineering reproductive sterility.

Ref: Giorda et al. (1990) Biochemistry 29: 7264-9
Tucker et al. (1988) Plant Physiol 88: 1257-62
Shani et al. (1997) 43: 837-42

Members of the family 20 glycosyl hydrolases catalyze the hydrolysis of terminal non-reducing N-acetyl-D-hexosamine residues in N-acetyl- β -D-hexosaminides. N-acetyl- β -glucosaminidase belongs to this family and exists in several different forms (consisting of various combinations of alpha and beta chains) depending on the organism. Family 20 glycosyl hydrolases have been implicated in lysosomal storage diseases (such as Sandhoff disease) and glycogen storage disease in humans. These types of proteins are also responsible for the hydrolysis of chitin. In plants, these proteins could be useful in controlling carbohydrate catabolism, thereby influencing the amount of sugars available for storage and/or use in other metabolic pathways. In addition, it is possible that such proteins could be used to engineer an endogenous insect protection mechanism, e.g. by secretion of a chitin-hydrolyzing composition by the plant.

Ref: Graham et al (1988) J Biol Chem 263: 16823-9
O'Dowd et al. (1988) Biochemistry 27: 5216-26

K. HMG box

The HMG box is a novel type of DNA-binding domain found in a diverse group of proteins. Numerous plant proteins contain this domain, such as the HMG1/2-like proteins. The expression of some of these HMG proteins appears to be regulated by circadian rhythms and in a light dependent manner, occurring at higher levels in roots, for example and lower levels in light-grown tissues such as cotyledons. Generally, HMG proteins are thought to influence transcription regulation. In plants, HMGs are believed to have a role in maintaining patterns of circadian-regulated expression for other genes, suggesting that these proteins could be exploited to control growth and development.

Ref: Laudet et al. (1993) Nucleic Acids Res 21: 2493-501
Zheng et al. (1993) Plant Mol Biol 23: 813-23
Grasser et al. (1993) Plant Mol Biol 23: 619-25

L. IL2

Interleukin-2 (IL-2) is produced in mammals by T cells in response to antigenic or mitogenic stimulation and is crucial for proper regulation and functioning of the immune response. IL-2 is capable of stimulating B cells, monocytes, lymphokine-activated killer cells, natural killer cells and glioma cells. Plant extracts have also been shown to stimulate the immune system (for example, mistletoe therapy for human cancer). It is known that IL-2 is involved in feedback inhibition pathways that impact the inflammatory response as well as the growth inhibition of tumor reactive T cells. Plant proteins containing IL-2-like sequences are useful as immunity-based therapeutics, acting in a manner similar to IL-2 in mammals.

Ref: Heike et al. (1997) Scand J Immunol 45: 221-6
Ariel et al. (1998) J Immunol 161: 2465-72
Schink (1997) Anticancer Drugs 8 Suppl 1: S47-51

M. Oxidored_FMN

NADPH dehydrogenases catalyze the reaction $\text{NADPH} + \text{acceptor} = \text{NADP}(+) + \text{reduced acceptor}$. One member of this family is yeast "old yellow enzyme" (OYE) and is thought to be involved in oxylipin metabolism. A second yeast family member is a protein that binds estrogen binding protein (EBP) in addition to exhibiting oxidoreductase activity. An Arabidopsis homolog to OYE has been described and estrogen binding proteins in plants have been reported. Plant proteins from this class have the potential to be used to modify lipid metabolism/catabolism. These proteins may also have use as therapeutics for breast and prostate cancer, and other abnormal growth in steroid-sensitive tissues.

Ref: Baker et al. (1998) Proc Soc Exp Biol Med 217: 317-21
Schaller and Weiler (1997) J Biol Chem 272: 28066-72
Mandani et al. (1994) PNAS USA 91: 922-6

N. Oxidored_q2

The NADH-plastoquinone oxidoreductases catalyze the reaction $\text{NADH} + \text{plastoquinone} = \text{NAD}(+) + \text{plastoquinol}$. In plants these reactions occur in the chloroplast and are believed to participate in a chloroplast respiratory system. Here, the NDH complex is postulated to act as

[illegible]

5 Kofer et al (1998) Mol Gen Genet 258: 166-73
Maier et al. (1995) J Mol Biol 251: 614-28

10 Polyadenylate binding proteins bind the poly (A) tail of mRNA. Plants, as exemplified by Arabidopsis, contain numerous PABP genes that are expressed in an organ-specific manner. For example, PABP2 is functional in roots and shoots, while PABP5 is expressed predominantly in immature flowers. The PABP proteins are implicated in numerous aspects of posttranscriptional regulation including mRNA turnover and translational initiation.

15 Control of activity of PABP proteins provides the ability to control the expression of various genes in particular organs during development.

20

25

30

Ref: Liu et al. (1997) J Gen Virol 78: 1265-70
Rohde et al. (1990) Virology 176: 648-51

Q. Pkinase_C

Plant serine/threonine protein kinases possessing this domain are expressed in all tissues and are known to undergo serine-specific autophosphorylation and specifically phosphorylate two ribosomal proteins, P14 and P16. During development, these proteins predominate during high metabolic activity in growing buds, root tips, leaf margins and germinating seeds. They are thought to be involved in the control of plant growth and development. In addition, two genes encoding proteins from this family have been described that help plant cells adapt during cold or high salt stresses. Consequently, engineering Pkinase C proteins provides a way to control general growth/development of the plant as well as a means to provide endogenous protection against environmental stresses.

Ref: Zhang et al. (1994) J Biol Chem 269: 17586-92

Mizoguchi et al. (1995) FEBS Lett 358: 199-204

R. REV

The REV proteins act post-transcriptionally to relieve negative repression of GAG and ENV production in retroviruses such as Human Immunodeficiency Virus type I (HIV-1). Plants contain retrovirus-like viruses such as pararetroviruses and retrotransposons (i.e. transposons having long terminal repeats). Plant retrotransposons in particular have been used to create mutations at various loci, thereby permitting gene isolation, gene tagging and the like. Manipulation of plant REV proteins enables control of transposition frequencies of corresponding transposable elements and provides a new tool for genetic engineering of plants.

Ref: Sodroski et al. (1986) Nature 321: 412-7

Franchini et al. (1989) PNAS USA 86: 2433-7

Marquet et al. (1995) 77: 113-24

Grandbastien et al. (1989) Nature 337: 376-80

Wright and Voytas (1998) Genetics 149: 703-15

S. RuBisCo small

Ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCo) catalyzes the initial step in the C3 photosynthetic carbon reduction cycle, adding carbon dioxide to D-ribulose 1,5-bisphosphate to form two molecules of 3-phospho-D-glycerate. RuBisCo is comprised of two subunits, one large which is synthesized in the chloroplast, and one small which is synthesized in the cytoplasm and then transported in to the chloroplast. The expression of the small subunit of RuBisCo is light regulated. Manipulation of these proteins could increase the efficiency of photosynthesis or allow alterations in developmental timing.

Ref: Giuliano et al. (1988) PNAS USA 85: 7089-93

Dedonder et al. (1993) Plant Physiol 101: 801-8

T. Sialyltransf

Members of the CMP-N-acetylneuraminate- β -galactosamide- α -2,3-sialyltransferase family catalyze the following reaction:

CMP-N-acetylneuraminate + β -D-galactosyl-1,3-N-acetyl- α -D-galactosaminyl-R = CMP + α -N-acetylneraminyl-2,3- β -D-galactosyl-1,3-N-acetyl- α -D-galactosaminyl-R. These proteins are thought to be responsible for the synthesis of the sequence neurac- α -2,3-gal- β -1,3-galnac- found on sugar chains)-linked to threonine or serine and also as a terminal sequence on certain gangliosides in mammalian cells. In plants, glycosyltransferases in the Golgi apparatus synthesize cell wall polysaccharides and elaborate the complex glycans of glycoproteins. Engineering of plant sialyltransferases allows targeting of proteins to particular cellular locations or enables the making of changes in cell wall structure.

Ref: Wee et al. (1998) Plant Cell 10: 1759-68

Lee et al. (1994) J Biol Chem 269: 10028-33

Kitagawa and Paulson (1994) J Biol Chem 269: 1394-401

U. Signal

Many plant proteins in this family contain sequences similar to those found in both components of the prokaryotic family of signal transducers known as the two-component systems. This suggests that activation may require a transfer of a phosphate group between

the transmitter domain and the receiver domain. One family member in Arabidopsis appears to be involved in ethylene (a plant hormone) signal transduction. Other proteins in this family appear to be involved in the regulation of gene transcription under conditions of environmental stress. Signal proteins can be exploited to affect plant growth and development and/or control plant responses to stress conditions such as cold, nutrient availability, etc.

Ref: Chang et al. (1993) Science 262: 539-44
Nagaya et al. (1993) Gene 131: 119-124
Gottfert et al. (1990) PNAS USA 87: 2680-4

V. vMSA

vMSA proteins are major surface antigens presenting on the envelope of various retroviruses. Surface antigens of retroviruses are often involved in tropism of the virus. Plants contain retrovirus-like viruses such as pararetroviruses and retrotransposons (i.e. transposons having long terminal repeats). Plant retrotransposons in particular have been used to create mutants at various loci, thereby permitting gene isolation, gene tagging and the like. Manipulation of plant vMSA proteins enables control of tropism of plant retroviruses that might be used for genetic engineering tools, thus enabling targeting of the virus to particular species and/or tissues of plants.

Ref: Okamoto et al. (1988) J Gen Virol 69: 2575-83
Grandbastien et al. (1989) Nature 337: 376-80
Wright and Voytas (1998) Genetics 149: 703-15

W. zf-CCCH

This family of proteins is defined by having two CX(8)CX(5)CX(3)H-type zinc finger domains. These proteins cover a broad range of functions. For example, the COP1 protein acts as a repressor of photomorphogenesis in darkness; light stimuli abolish this suppressive action. In addition, COP1 protein can function as a negative transcriptional regulator capable of direct interaction with components of the G-protein signaling pathway. As a second example, a zf-CCCH protein identified in Arabidopsis appears to be involved in the resistance to DNA damage induced by UV light and chemical DNA-damaging agents.

Overexpression of this class of proteins permits production of plants that are better suited to adverse environments. Manipulation of expression of zf-CCCH proteins functioning as transcriptional regulators, such as COP1, enables manipulation of some signal transduction pathways.

Ref: Pang et al. (1993) Nucleic Acids Res 21: 1647-53
Deng et al. (1992) Cell 71: 791-801

X. zf-RanBP

Proteins falling within this category contain many X-X-F-G and X-F-X-F-G repeats, and may contain RANBP1-like or PPIase domains. Plant proteins having domains similar to these include PAS1 and GMSTI. PAS1 has been shown to have dramatic developmental affects that appear to be correlated with both cell division and cell wall elongation. GMSTI has high identity to the yeast STI stress-inducible gene and has been shown to be heat inducible. Proteins such as these may be useful for controlling growth and form of development.

Ref: Vittorioso et al. (1998) Mol Cell Biol 18: 3034-43
Hernandez Torres et al. (1995) 27: 1221-6

Y. Peptidase M48.

Proteins belonging to this peptidase family are metalloproteases that bind zinc as a cofactor and are located in the membranes of the endoplasmic reticulum. They function in NH₂-terminal proteolytic processing, as shown for the yeast STE24 gene product. This gene is required for the correct processing of α -factor, a yeast pheromone. Family M48 peptidases also appear to be required for some prenylation reactions, mediating COOH-terminal CAAX processing. Prenylation reactions are believed to be involved in the regulation of protein-protein and protein-membrane interactions. As an example, RAS GTPase activity is regulated in part by localization to the inner side of the plasma membrane upon prenylation. In plants, proteins from this family could be involved in pollen-stigma interactions such as those mediating self-pollination vs. outcrossing, or could be members of several secondary metabolism pathways.

Ref: Fujimura-Kamada et al. (1997) J Cell Biol. 136: 271-85. Tam et al. (1998) J Cell Biol. 142: 635-49.

Z. DNA Pol Viral N

The DNA pol Viral N domain is located at the N-terminal region of DNA polymerase isolated from several retroviral viruses such as the Cauliflower Mosaic Virus. The domain motif has also been found in numerous other species from humans to cyanobacteria. In these organisms, this motif seems to be associated with two types of sequences; retrotransposons and mitochondrial genes. In the mitochondrial sequences this domain is potentially involved in the self-splicing conducted by group II introns. Various manipulations of this gene in plants allows control of the numerous retrotransposons endogenous to plant genomes or allows engineering of mitochondrial function, especially to increase efficiency of energy utilization by cells.

REF: Chapdelaine and Bonen (1991) Cell 65: 465-72

Ferat and Miche (1993) Nature 364: 358-61

Wilson et al. (1994) 368: 32-8

Cambareri et al. (1994) 242: 658-65

Gaardner et al. (1981) NAR 9: 2871-2888

Cummings et al. (1990) Curr Genet 17: 375-402

Hattori et al. (1986) Nature 321: 625-8

Aa. Calpain inhib

This domain is found in calpastatin, an inhibitor protein specific for calpain. Calpain is a non-lysosomal calcium-dependent intracellular protease that appears to be involved in the dynamic changes of the cytoskeleton, especially actin-related structures, during early *Drosophila* embryogenesis [1]. Calpastatins co-exist in cells with calpains and the subcellular distribution of calpastatin is thought to be important to calpain regulation [2]. In plants calpains and calpastatins could be involved in embryogenesis and non-embryogenic organ reiteration. Mutations occurring in calpain inhibitor repeat domains would produce developmental abnormalities such as abnormal leaf, root or flower development.

Refs

- 1 Emori Y and Saigo K (1994) J Biol Chem 269: 25137-42.
- 2 Mellgren RL, Lane RD, Mericle MT (1989) Biochim Biophys Acta 999: 71-77.

Ab. chorismate_bind

5 Chorismate binding domains are present in plant anthranilate synthase (AS) genes. AS genes catalyze the first step in the biosynthesis of tryptophan by converting chorismate and L-glutamine to anthranilate, pyruvate and L-glutamate. Some of these genes are involved in feedback inhibition by tryptophan [1] while some are feedback insensitive [2]. In Arabidopsis, two AS genes have overlapping, but different distributions. One of these AS
10 genes is induced by wounding and bacterial pathogen infiltration [1]. Mutations in the chorismate binding domain would affect the production of tryptophan and could influence the plant's defense system. AS gene products can be used for *in vitro* synthesis of tryptophan and tryptophan derivatives.

15 Refs

- 1 Niyogi KK, Fink GR (1992) Plant Cell 4: 721-33.
- 2 Song HS, Brotherton JE, Gonzales RA, Wilholm JM (1998) Plant Physiol 117:533-43.

20 Ac. late_protein_L2

Papillomaviruses are encapsulated double stranded DNA viruses. Plants are susceptible to infection by double stranded DNA viruses such as Cauliflower Mosaic virus (CaMV). The coat proteins in these plant viruses are critical to the virus life cycle within the plant. For example, the coat protein of CaMV is thought to be involved in intra- and inter-cellular
25 movement within the plant [1]. Engineering of proteins having similarity to papillomavirus coat proteins may enable the production of plants having better resistance to natural plant double stranded DNA viruses.

Refs

- 30 1 Thompson SR, Melcher U (1993) J Gen Virol 74: 1141-8.

Ad. Peptidase_M41

Proteins belonging to this peptidase family are metalloproteases that bind zinc as a cofactor and are integral membrane proteins. They seem to be involved in the degradation of carboxy-

09669960 1013300

terminal-tagged cytoplasmic proteins. In plants, these proteins are located in the thylakoid membranes of the chloroplasts, their expression is light regulated and they are thought to be involved in degradation of soluble stromal proteins and turn-over of thylakoid proteins [1].

Manipulation of expression and structure of these proteins would have effects on the efficiency of photosynthesis and the development of chloroplasts.

Refs

1 Lindahl M, Tabak s, Cseke L, Pichersky E, Andersson B, Adam Z (1996) J Biol Chem 271: 29329-34.

Ae. UPF0051

There is some evidence that, in plants, proteins in this family are involved in ATP synthesis in chloroplasts [1, 2]. Mutations in these proteins or altering their expression would affect the efficiency of photosynthesis and energy production.

Refs

1 Kostrzewa M, Zetsche K (1992) J Mol Biol 227: 961-70.

2 Kostrzewa M, Zetsche K (1993) Plant Mol Biol 23: 67-76

Af. E7

Papillomaviruses are encapsulated double stranded DNA viruses. The Papillomavirus early protein 7 (E7) is known as a potent immortalizing and transforming agent. Transformation by E7 is thought to be mediated by the physical association of E7 with cellular proteins regulating entry into the cell cycle [1]. The result is entry into the cell cycle and suppression of terminal differentiation in mammalian cells. Thus, engineering of proteins having similarity to papillomavirus E7 protein enables the production of plants having altered cellular proliferation characteristics and possibly altered morphology. For example, overexpression of E7-like proteins would be expected to result in proliferation of cells of the tissue in which the E7 protein is expressed, perhaps with suppression of differentiation events. Thus, for example, overexpression of E7-like proteins in meristem cells can result in taller plants and suppression of leafing and/or flowering.

Refs

1 Zwerschke W, Jansen-Durr P Adv Cancer Res 2000;78:1-29

Ag. Peptidase U7

This protein is known to be an integral membrane protein in the cyanobacterium Synechocystis where it functions to digest cleaved signal peptides [1]. This activity is necessary to maintain proper secretion of mature proteins across the membrane. In higher plants this protein may be present in the plastid or chloroplast membranes where it would function by enabling protein movement into and out of the chloroplasts. Mutations in this protein would be expected to affect the development of plastids, including chloroplasts, or alter the energy transfer system within the chloroplasts, thereby affecting growth and development.

Refs

- 1 Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirose M, Sugiura M, Sasamoto S, Kimura T, Hosouchi T, Matsuno A, Muraki A, Nakazaki N, Naruo K, Okumura S, Shimpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M, Tabata S (1996) DNA Res 3:109-36.

Ah. 5'-3' Exonuclease

The 5'-3' exonuclease domain is one found in bacterial DNA polymerases I and in yeast DNA repair enzymes such as Exonuclease I. Yeast Exo I is involved in mitotic recombination and also includes a domain that interacts with the mismatch repair protein MSH2. The 5'-3' exonuclease domain is also present in XPG DNA repair enzymes in humans and in yeast RAD9 protein. Defects in XPG proteins result in Xeroderma Pigmentosum. Thus defects in 5'-3' exonuclease domain-containing proteins in plants are expected to lead to defects in DNA repair and corresponding high spontaneous and inducible mutation rates. Consensus sequence:

IMKKKLLLVDGSSLAFFALPPLTNSAGEPTNAVYGFLLKMLIKLIEQEQPTHIAVV
FDAQAKTFRHELYEGYKAGRAP
TPDELREQIPLIKELLDALGIPLLVAGYEADDVIGTLAKLAEKEGYEVLIVTGDRDLL
QLVSDHVTVIITKKGIAEFTL
FTPEAVIEKYGLTPEQIIDYKALMGDSSDNIPGVKGIGEKTAACKLLQEYGSLEGIYANL
DKLKGKKLREKLLAHKEDAKL
SRDLATIKTDVPLDLTLDDLRLPDPDRDALDLLFDE

Pfam	Prosite	Full Name	Description
3_5_exonuclease		3'-5' exonuclease	<p>Accession number: PF01612</p> <p>Definition: 3'-5' exonuclease</p> <p>Author: Bashton M, Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_659 (release 4.1)</p> <p>Gathering cutoffs: -11 -11</p> <p>Trusted cutoffs: -10.70 -10.70</p> <p>Noise cutoffs: -24.50 -24.50</p> <p>HMM build command line: hmmbuild HMM SEED</p> <p>HMM build command line: hmmcalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 85137890</p> <p>Reference Title: Structure of large fragment of Escherichia coli DNA polymerase I complexed with dTMP.</p> <p>Reference Author: Ollis DL, Brick P, Hamlin R, Xuong NG, Steitz TA;</p> <p>Reference Location: Nature 1985;313:762-766.</p> <p>Reference Number: [2]</p> <p>Reference Medline: 98060913</p> <p>Reference Title: The proofreading domain of Escherichia coli DNA polymerase I and other DNA and/or RNA exonuclease domains.</p> <p>Reference Author: Moser MJ, Holley WR, Chatterjee A, Mian IS;</p> <p>Reference Location: Nucleic Acids Res 1997;25:5110-5118.</p> <p>Reference Number: [3]</p> <p>Reference Medline: 98361165</p> <p>Reference Title: Replication focus-forming activity 1 and the Werner syndrome gene product</p> <p>Reference Author: Yan H, Chen CY, Kobayashi R, Newport J;</p> <p>Reference Location: Nat Genet 1998;19:375-378.</p> <p>Reference Number: [4]</p> <p>Reference Medline: 97434221</p> <p>Reference Title: The Werner syndrome protein is a DNA helicase.</p> <p>Reference Author: Gray MD, Shen JC, Kamath-Loeb AS, Blank A, Sopher BL;</p> <p>Reference Author: Martin GM, Oshima J, Loeb LA;</p> <p>Reference Location: Nat Genet 1997;17:100-103.</p> <p>Reference Number: [5]</p> <p>Reference Medline: 97370026</p> <p>Reference Title: DNA helicase activity in Werner's syndrome gene product synthesized in a baculovirus system.</p> <p>Reference Author: Suzuki N, Shimamoto A, Imamura O, Kuromitsu J, Kitao S,</p> <p>Reference Author: Goto M, Furuichi Y;</p> <p>Reference Location: Nucleic Acids Res 1997;25:2973-2978.</p> <p>Database Reference: SCOP; 1dpi; fa; [SCOP-USA][CATH-PDBSUM]</p> <p>Database Reference: INTERPRO; IPR002562;</p> <p>Database Reference: PDB; 1kfd ; 348; 518;</p> <p>Database Reference: PDB; 1d8y A; 348; 518;</p> <p>Database Reference: PDB; 1d9d A; 348; 518;</p> <p>Database Reference: PDB; 1d9f A; 348; 518;</p> <p>Database Reference: PDB; 1kfs A; 348; 518;</p> <p>Database Reference: PDB; 1kln A; 348; 518;</p> <p>Database Reference: PDB; 1krp A; 348; 518;</p> <p>Database Reference: PDB; 1ksp A; 348; 518;</p> <p>Database Reference: PDB; 1qsl A; 348; 518;</p> <p>Database Reference: PDB; 2kfn A; 348; 518;</p> <p>Database Reference: PDB; 2kfz A; 348; 518;</p> <p>Database Reference: PDB; 2kzm A; 348; 518;</p> <p>Database Reference: PDB; 2kzz A; 348; 518;</p> <p>Comment: This domain is responsible for the 3'-5' exonuclease proofreading</p> <p>Comment: activity of E. coli DNA polymerase I (poll) and other enzymes,</p>

Comment: This domain is responsible for the 3'-5' exonuclease proofreading activity of E. coli DNA polymerase I (polI) and other enzymes.

Pfam	Prosite	Full Name	Description
			<p>Comment: it catalyses the hydrolysis of unpaired or mismatched nucleotides.</p> <p>Comment: This domain consists of the amino-terminal half of the Klenow fragment</p> <p>Comment: in E. coli polI it is also found in the Werner syndrome helicase</p> <p>Comment: (WRN), focus forming activity 1 protein (FFA-1) and ribonuclease D</p> <p>Comment: (RNase D).</p> <p>Comment: Werner syndrome is a human genetic disorder causing premature aging;</p> <p>Comment: the WRN protein has helicase activity in the 3'-5' direction [4,5].</p> <p>Comment: The FFA-1 protein is required for formation of a replication foci</p> <p>Comment: and also has helicase activity; it is a homologue of the WRN</p> <p>Comment: protein [3].</p> <p>Comment: RNase D is a 3'-5' exonuclease involved in tRNA processing.</p> <p>Comment: Also found in this family is the autoantigen PM/Scl thought to be</p> <p>Comment: involved in polymyositis-scleroderma overlap syndrome.</p> <p>Number of members: 41</p>
3HCDH	PDOC00065	3-hydroxyacyl-CoA dehydrogenase signature	<p>3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35) (HCDH) [1] is an enzyme involved in fatty acid metabolism, it catalyzes the reduction of 3-hydroxyacyl-CoA to 3-oxoacyl-CoA. Most eukaryotic cells have 2 fatty-acid beta-oxidation systems, one located in mitochondria and the other in peroxisomes. In peroxisomes 3-hydroxyacyl-CoA dehydrogenase forms, with enoyl-CoA hydratase (ECH) and 3,2-trans-enoyl-CoA isomerase (ECI) a multifunctional enzyme where the N-terminal domain bears the hydratase/isomerase activities and the C-terminal domain the dehydrogenase activity. There are two mitochondrial enzymes: one which is monofunctional and the other which is, like its peroxisomal counterpart, multifunctional.</p> <p>In Escherichia coli (gene fadB) and Pseudomonas fragi (gene faoA) HCDH is part of a multifunctional enzyme which also contains an ECH/ECI domain as well as a 3-hydroxybutyryl-CoA epimerase domain [2].</p> <p>The other proteins structurally related to HCDH are:</p> <ul style="list-style-type: none"> - Bacterial 3-hydroxybutyryl-CoA dehydrogenase (EC 1.1.1.157) which reduces 3-hydroxybutanoyl-CoA to acetoacetyl-CoA [3]. - Eye lens protein lambda-crystallin [4], which is specific to lagomorphes (such as rabbit). <p>There are two major region of similarities in the sequences of proteins of the HCDH family, the first one located in the N-terminal, corresponds to the NAD-binding site, the second one is located in the center of the sequence. We have chosen to derive a signature pattern from this central region.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [DNE]-x(2)-[GA]-F-[LIVMFY]-x-[NT]-R-x(3)-</p>

Pfam	Prosite	Full Name	Description
			<p>[PA]-[LIVMFY](2)-x(5)-[LIVMFYCT]-[LIVMFY]-x(2)-[GV] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update July 1998 / Pattern and text revised. References [1] Birktoff J.J., Holden H.M., Hamlin R., Xuong N.-H., Banaszak L.J. Proc. Natl. Acad. Sci. U.S.A. 84:8262-8266(1987). [2] Nakahigashi K., Inokuchi H. Nucleic Acids Res. 18:4937-4937(1990). [3] Mullany P., Clayton C.L., Pallen M.J., Slone R., Al-Saleh A., Tabaqchali S. FEMS Microbiol. Lett. 124:61-67(1994). [4] Mulders J.W.M., Hendriks W., Blankesteyn W.M., Bloemendal H., de Jong W.W. J. Biol. Chem. 263:15462-15466(1988).</p>
4HPPD_C		4-hydroxyphenylpyruvate dioxxygenase C terminal domain	<p>Accession number: PF01626 Definition: 4-hydroxyphenylpyruvate dioxxygenase C terminal domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1116 (release 4.1) Gathering cutoffs: -35 -35 Trusted cutoffs: -25.80 -25.80 Noise cutoffs: -44.90 -44.90 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 93279307 Reference Title: Human 4-hydroxyphenylpyruvate dioxxygenase. Primary Reference Title: structure and chromosomal localization of the gene. Reference Author: Ruetschi U, Dellsen A, Sahlin P, Stenman G, Rymo L, Reference Author: Lindstedt S; Reference Location: Eur J Biochem 1993;213:1081-1089. Database Reference: INTERPRO; IPR002887; Comment: 4-Hydroxyphenylpyruvic acid dioxxygenase (HPD) is an important enzyme Comment: in tyrosine catabolism in most organisms. A genetic deficiency in Comment: this enzyme in humans and mice leads to hereditary tyrosinemia type 3. Comment: The identity of the C-terminus of the HPD makes this part of the Comment: molecule a candidate for a functional role in the catalytic process Comment: [1]. This region is found as a separate protein Swiss:Q49717 that Comment: is somewhat different from HPD and may have a different but related Comment: protein function (Unpublished observation Bateman A). Number of members: 28</p>

Pfam	Prosite	Full Name	Description
5_3_exonuclease		5'-3' exonuclease domain	<p>The 5'-3' exonuclease domain is one found in bacterial DNA polymerases I and in yeast DNA repair enzymes such as Exonuclease I. Yeast Exo I is involved in mitotic recombination and also includes a domain that interacts with the mismatch repair protein MSH2. The 5'-3' exonuclease domain is also present in XPG DNA repair enzymes in humans and in yeast RAD9 protein. Defects in XPG proteins result in Xeroderma Pigmentosum. Thus defects in 5'-3' exonuclease domain-containing proteins in plants are expected to lead to defects in DNA repair and corresponding high spontaneous and inducible mutation rates. Consensus sequence:</p> <p>IMKKKLLLVGSSLAFFRAFFALPPLTNSAGEPTNAVYGFCLKMLIK LIEQEQPTHIAVVFDAKAKTFRHELYEGYKAGRAP TPDELREQIPLIKELLDALGIPLLEVAGYEADDVIGTLAKLAEKEG YEVLI VTGDRDLLQLVSDHVTIITKKGIAEFTL FTPEAVIEKYGLTPEQIIDYKALMGDSSDNIPGVKGIGEKTAAKLL QEYGSLEGYIANLDKLGKGLREKLLAHKEDAKL SRDLATIKTDVPLDLTLDDLRLPDPDRDALDLLFDE</p> <p>Ref: Fiorentini P. et al. RT. Mol. Cell. Biol. 17:2764-2773(1997). Tishkoff et al. Cancer Res. 0:0-0(1998). Macinnes M.A. et al. Mol. Cell. Biol. 13:6393-6402(1993).</p>
60s_ribosomal		60s Acidic ribosomal protein	<p>Accession number: PF00428 Definition: 60s Acidic ribosomal protein Author: Finn RD Alignment method of seed: Clustalw Source of seed members: Pfam-B_151 (release 1.0) Gathering cutoffs: 17 17 Trusted cutoffs: 17.80 17.80 Noise cutoffs: 9.30 9.30 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96282699 Reference Title: Proteins P1, P2, and P0, components of the eukaryotic Reference Title: ribosome stalk. New structural and functional aspects. Reference Author: Remacha M, Jimenez-Diaz A, Santos C, Briones E, Zambrano R, Reference Author: Rodriguez Gabriel MA, Guarinos E, Ballesta JP; Reference Location: Biochem Cell Biol 1995;73:959-968. Database Reference: INTERPRO; IPR001813; Database reference: PFAMB; PB002218; Comment: This family includes archaeobacterial L12, eukaryotic P0, P1 and P2. Number of members: 109</p>
6PF2K	PDOC00158	Phosphoglycerate mutase family phosphohistidine signature	<p>Phosphoglycerate mutase (EC 5.4.2.1) (PGAM) and bisphosphoglycerate mutase (EC 5.4.2.4) (BPGM) are structurally related enzymes which catalyze reactions involving the transfer of phospho groups between the three carbon atoms of phosphoglycerate [1,2]. Both enzymes can catalyze three different reactions, although in different proportions:</p> <ul style="list-style-type: none"> - The isomerization of 2-phosphoglycerate (2-PGA) to 3-phosphoglycerate (3-PGA) with 2,3-diphosphoglycerate (2,3-DPG) as the primer of the reaction. - The synthesis of 2,3-DPG from 1,3-DPG with 3-PGA as a primer. - The degradation of 2,3-DPG to 3-PGA (phosphatase EC 3.1.3.13 activity). <p>In mammals, PGAM is a dimeric protein. There are two isoforms of PGAM: the M</p>

0969980 101300

Pfam	Prosite	Full Name	Description
			<p>(muscle) and B (brain) forms. In yeast, PGAM is a tetrameric protein. BPGM is a dimeric protein and is found mainly in erythrocytes where it plays a major role in regulating hemoglobin oxygen affinity as a consequence of controlling 2,3-DPG concentration.</p> <p>The catalytic mechanism of both PGAM and BPGM involves the formation of a phosphohistidine intermediate [3].</p> <p>The bifunctional enzyme 6-phosphofructo-2-kinase / fructose-2,6-bisphosphatase (EC 2.7.1.105 and EC 3.1.3.46) (PF2K) [4] catalyzes both the synthesis and the degradation of fructose-2,6-bisphosphate. PF2K is an important enzyme in the regulation of hepatic carbohydrate metabolism. Like PGAM/BPGM, the fructose-2,6-bisphosphatase reaction involves a phosphohistidine intermediate and the phosphatase domain of PF2K is structurally related to PGAM/BPGM.</p> <p>The bacterial enzyme alpha-ribazole-5'-phosphate phosphatase (gene cobC) which is involved in cobalamin biosynthesis also belongs to this family [5].</p> <p>We built a signature pattern around the phosphohistidine residue.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [LIVM]-x-R-H-G-[EQ]-x(3)-N [H is the phosphohistidine residue] Sequences known to belong to this class detected by the pattern ALL, except for Haemophilus influenzae PGAM. Other sequence(s) detected in SWISS-PROT 2.</p> <p>Note some organisms harbor a form of PGAM independent of 2,3-DPG, this enzyme is not related to the family described above [6]. Last update November 1995 / Text revised. References [1] Le Boulch P., Joulin V., Garel M.-C., Rosa J., Cohen-Solal M. Biochem. Biophys. Res. Commun. 156:874-881(1988).</p> <p>[2] White M.F., Fothergill-Gilmore L.A. FEBS Lett. 229:383-387(1988).</p> <p>[3] Rose Z.B. Meth. Enzymol. 87:43-51(1982).</p> <p>[4] Bazan J.F., Fletterick R.J., Pilgis S.J. Proc. Natl. Acad. Sci. U.S.A. 86:9642-9646(1989).</p> <p>[5] O'Toole G.A., Trzebiatowski J.R., Escalante-Semerena J.C. J. Biol. Chem. 269:26503-26511(1994).</p> <p>[6] Grana X., De Lecea L., El-Maghrabi M.R., Urena J.M., Caellas C., Carreras J., Puigdomenech P., Pilgis S.J., Climent F. J. Biol. Chem. 267:12797-12803(1992).</p>
7tm_5		7TM chemoreceptor	<p>Accession number: PF01604 Definition: 7TM chemoreceptor</p>

Pfam	Prosite	Full Name	Description
			<p>Author: Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_942 (release 4.1)</p> <p>Gathering cutoffs: -46 -46</p> <p>Trusted cutoffs: -44.30 -44.30</p> <p>Noise cutoffs: -47.80 -47.80</p> <p>HMM build command line: hmmbuild -F HMM SEED</p> <p>HMM build command line: hmmscalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 98248686</p> <p>Reference Title: Two large families of chemoreceptor genes in the nematodes</p> <p>Reference Title: Caenorhabditis elegans and Caenorhabditis briggsae reveal</p> <p>Reference Title: extensive gene duplication, diversification, movement, and</p> <p>Reference Title: intron loss.</p> <p>Reference Author: Robertson HM;</p> <p>Reference Location: Genome Res 1998;8:449-463.</p> <p>Database Reference: INTERPRO; IPR003003;</p> <p>Comment: This large family of proteins are related to 7tm_1.</p> <p>Comment: They are 7 transmembrane receptors. This family does not</p> <p>Comment: include all known members, as there are problems with</p> <p>Comment: overlapping specificity with 7tm_1.</p> <p>Comment: This family is greatly expanded in the nematode worm C.</p> <p>Comment: elegans.</p> <p>Number of members: 180</p>
Aa_trans		Transmembrane amino acid transporter protein	<p>Accession number: PF01490</p> <p>Definition: Transmembrane amino acid transporter protein</p> <p>Author: Bashton M, Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_419 (release 4.0)</p> <p>Gathering cutoffs: 25 25</p> <p>Trusted cutoffs: 150.80 150.80</p> <p>Noise cutoffs: 3.60 3.60</p> <p>HMM build command line: hmmbuild -F HMM SEED</p> <p>HMM build command line: hmmscalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 98007977</p> <p>Reference Title: Identification and characterization of the vesicular GABA</p> <p>Reference Title: transporter.</p> <p>Reference Author: McIntire SL, Reimer RJ, Schuske K, Edwards RH, Jorgensen</p> <p>Reference Author: EM;</p> <p>Reference Location: Nature 1997;389:870-876.</p> <p>Database Reference: INTERPRO; IPR002422;</p> <p>Database reference: PFAMB; PB020912;</p> <p>Comment: This transmembrane region is found in many amino acid transporters</p> <p>Comment: including UNC-47 and MTR. UNC-47 encodes a vesicular amino butyric acid</p> <p>Comment: (GABA) transporter, (VGAT). UNC-47 is predicted to have 10 transmembrane</p> <p>Comment: domains Swiss:P34579 [1]. MTR is a N system amino acid transporter system</p> <p>Comment: protein involved in methyltryptophan resistance Swiss:P38680.</p> <p>Comment: Other members of this family include proline transporters and amino</p> <p>Comment: acid permeases.</p> <p>Number of members: 50</p>
ABC_tran	PDOC00185	ABC transporters family signature	<p>On the basis of sequence similarities a family of related ATP-binding proteins has been characterized [1 to 5]. These proteins are associated with a variety of distinct biological processes in both prokaryotes and</p>

Pfam	Prosite	Full Name	Description
			<p>eukaryotes, but a majority of them are involved in active transport of small hydrophilic molecules across the cytoplasmic membrane. All these proteins share a conserved domain of some two hundred amino acid residues, which includes an ATP-binding site. These proteins are collectively known as ABC transporters. Proteins known to belong to this family are listed below (references are only provided for recently determined sequences).</p> <p>In prokaryotes:</p> <ul style="list-style-type: none"> - Active transport systems components: alkylphosphonate uptake(phnC/phnK/phnL); arabinose (araG); arginine (artP); dipeptide (dciAD;dppD/dppF); ferric enterobactin (fepC); ferrichrome (fhuC); galactoside (mglA); glutamine (glnQ); glycerol-3-phosphate (ugpC); glycine betaine/L-proline (proV); glutamate/aspartate (gltL); histidine (hisP); iron(III) (sfuC); iron(III) dicitrate (fecE); lactose (lack); leucine/isoleucine/valine (braF/braG;livF/livG); maltose (malk); molybdenum (modC); nickel (nikD/nikE); oligopeptide (amiE/amiF;oppD/oppF); peptide (sapD/sapF); phosphate (pstB); putrescine (potG); ribose (rbsA); spermidine/putrescine (potA); sulfate (cysA); vitamin B12 (btuD). - Hemolysin/leukotoxin export proteins hlyB, cyaB and lktB. - Colicin V export protein cvaB. - Lactococcin export protein lcnC [6]. - Lantibiotic transport proteins nisT (nisin) and spaT (subtilin). - Extracellular proteases B and C export protein prtD. - Alkaline protease secretion protein aprD. - Beta-(1,2)-glucan export proteins chvA and ndvA. - Haemophilus influenzae capsule-polysaccharide export protein bexA. - Cytochrome c biogenesis proteins ccmA (also known as cycV and helA). - Polysialic acid transport protein kpsT. - Cell division associated ftsE protein (function unknown). - Copper processing protein nosF from Pseudomonas stutzeri. - Nodulation protein nodI from Rhizobium (function unknown). - Escherichia coli proteins cydC and cydD. - Subunit A of the ABC excision nuclease (gene uvrA). - Erythromycin resistance protein from Staphylococcus epidermidis (gene msrA). - Tylosin resistance protein from Streptomyces fradiae (gene tlrC) [7]. - Heterocyst differentiation protein (gene hetA) from Anabaena PCC 7120. - Protein P29 from Mycoplasma hyorhinis, a probable component of a high affinity transport system. - yhbG, a putative protein whose gene is linked with ntrA in many bacteria such as Escherichia coli, Klebsiella pneumoniae, Pseudomonas putida, Rhizobium meliloti and Thiobacillus ferrooxidans. - Escherichia coli and related bacteria hypothetical proteins yabJ, yadG, yagC, ybbA, ycjW, yddA, yehX, yejF, yheS, yhiG, yhiH, yjcW, yjjK, yojl, yrbF and ytfR. <p>In eukaryotes:</p>

0966980 "101300

Pfam	Prosite	Full Name	Description
			<p>- The multidrug transporters (Mdr) (P-glycoprotein), a family of closely related proteins which extrude a wide variety of drugs out of the cell (for a review see [8]).</p> <p>- Cystic fibrosis transmembrane conductance regulator (CFTR), which is most probably involved in the transport of chloride ions.</p> <p>- Antigen peptide transporters 1 (TAP1, PSF1, RING4, HAM-1, mtp1) and 2 (TAP2, PSF2, RING11, HAM-2, mtp2), which are involved in the transport of antigens from the cytoplasm to a membrane-bound compartment for association with MHC class I molecules.</p> <p>- 70 Kd peroxisomal membrane protein (PMP70).</p> <p>- ALDP, a peroxisomal protein involved in X-linked adrenoleukodystrophy [9].</p> <p>- Sulfonylurea receptor [10], a putative subunit of the B-cell ATP-sensitive potassium channel.</p> <p>- Drosophila proteins white (w) and brown (bw), which are involved in the import of ommatidium screening pigments.</p> <p>- Fungal elongation factor 3 (EF-3).</p> <p>- Yeast STE6 which is responsible for the export of the a-factor pheromone.</p> <p>- Yeast mitochondrial transporter ATM1.</p> <p>- Yeast MDL1 and MDL2.</p> <p>- Yeast SNQ2.</p> <p>- Yeast sporidesmin resistance protein (gene PDR5 or STS1 or YDR1).</p> <p>- Fission yeast heavy metal tolerance protein hmt1. This protein is probably involved in the transport of metal-bound phytochelatin.</p> <p>- Fission yeast brefeldin A resistance protein (gene bfr1 or hba2).</p> <p>- Fission yeast leptomycin B resistance protein (gene pmd1).</p> <p>- mbpX, a hypothetical chloroplast protein from Liverwort.</p> <p>- Prestalk-specific protein tagB from slime mold. This protein consists of two domains: a N-terminal subtilase catalytic domain (see <PDOC00125>) and a C-terminal ABC transporter domain.</p> <p>As a signature pattern for this class of proteins, we use a conserved region which is located between the 'A' and the 'B' motifs of the ATP-binding site.</p> <p style="text-align: right;">Consensus pattern</p> <p>[LIVMFYC]-[SA]-[SAPGLVFKQH]-G-[DENQMW]-[KRQASPLIMFW]-[KRNQSTAVM]-[KRACLVM]-[LIVMFYPAN]-{PHY}-[LIVMFW]-[SAGCLIVP]-{FYWHP}-[KRHP]-[LIVMFYWSTA] Sequences known to belong to this class detected by the pattern ALL, except for 25 sequences. Other sequence(s) detected in SWISS-PROT 42. Note the ATP-binding region is duplicated in araG, mdl, mdrA, rbsA, tlc, uvrA, yefF, Mdr's, CFTR, pmd1 and in EF-3. In some of those proteins, the above pattern only detect one of the two copies of the domain. Note the proteins belonging to this family also contain one or two copies of the ATP-binding motifs 'A' and 'B' (see <PDOC00017>).</p> <p style="text-align: right;">[1]</p> <p>Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P.</p> <p>J. Bioenerg. Biomembr. 22:571-592(1990).</p> <p>[2]</p> <p>Higgins C.F., Gallagher M.P., Mimmack M.M., Pearce S.R. BioEssays 8:111-116(1988).</p> <p>[3]</p> <p>Higgins C.F., Hiles I.D., Salmond G.P.C., Gill D.R., Downie J.A., Evans I.J., Holland I.B., Gray L., Buckels S.D., Bell A.W., Hermodson M.A.</p> <p>Nature 323:448-450(1986).</p>

Pfam	Prosite	Full Name	Description
			<p>[4] Doolittle R.F., Johnson M.S., Husain I., van Houten B., Thomas D.C., Sancar A. <i>Nature</i> 323:451-453(1986).</p> <p>[5] Blight M.A., Holland I.B. <i>Mol. Microbiol.</i> 4:873-880(1990).</p> <p>[6] Stoddard G.W., Petzel J.P., van Belkum M.J., Kok J., McKay L.L. <i>Appl. Environ. Microbiol.</i> 58:1952-1961(1992).</p> <p>[7] Rosteck P.R. Jr., Reynolds P.A., Hershberger C.L. <i>Gene</i> 102:27-32(1991).</p> <p>[8] Gottesman M.M., Pastan I. <i>J. Biol. Chem.</i> 263:12163-12166(1988).</p> <p>[9] Valle D., Gaertner J. <i>Nature</i> 361:682-683(1993).</p> <p>[10] Aguilar-Bryan L., Nichols C.G., Wechsler S.W., Clement J.P. IV, Boyd A.E. III, Gonzalez G., Herrera-Sosa H., Nguy K., Bryan J., Nelson D.A. <i>Science</i> 268:423-426(1995).</p>
ABC2_membrane	PDOC00692	ABC-2 type transport system integral membrane proteins signature	<p>Integral membrane components of a number of bacterial active transport systems have been shown to be evolutionary related and to form a distinct family [1,2]. These proteins are:</p> <ul style="list-style-type: none"> - <i>Escherichia coli</i> kpsM, involved in polysialic acid export. - <i>Haemophilus influenzae</i> bexB, involved in polyribosylribitol phosphate capsule polysaccharide export. - <i>Salmonella typhi</i> vexB, involved in translocation of the Vi polysaccharide. - <i>Neisseria meningitidis</i> ctrC, involved in polyneuraminic acid capsule polysaccharide export. - <i>Rhizobiaceae</i> nodulation protein J (gene nodJ), probably involved in exporting a modified beta-1,4-linked N-acetylglucosamine oligosaccharide. - <i>Streptomyces peucetius</i> drrB, involved in exporting the antibiotics daunorubicin and doxorubicin. - <i>Klebsiella pneumoniae</i> O-antigen exprt system protein rfbA. - <i>Yersinia enterocolitica</i> O-antigen exprt system protein rfbD. - <i>Escherichia coli</i> hypothetical protein yadH. - <i>Escherichia coli</i> hypothetical protein yhhJ. <p>The molecular size of these proteins is around 30 Kd. They are thought to contain six transmembrane regions. They either form homooligomeric channels or associate with another type of transmembrane protein to form heterooligomers.</p> <p>Transport systems in which they participate are energized by an ATP-binding protein that belongs to the ABC transporter family. The designation 'ABC-2' has been proposed [1] for these transport systems.</p> <p>As a signature pattern, we selected a conserved region located in the C-terminal section of these proteins.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [LIMST]-x(2)-[LIMW]-x(2)-[LIMCA]-[GSTC]-x-[GSAIV]-x(6)-[LIMGA]-[PGSNQ]-x(9,12)-P-[LIMFT]-x-[HRSY]-</p>

Pfam	Prosite	Full Name	Description
			<p>x(5)-[RQ] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 2. Last update November 1997 / Pattern and text revised. References [1] Reizer J., Reizer A., Saier M.H. Jr. Protein Sci. 1:1326-1332(1992).</p> <p>[2] Vazquez M., Santana O., Quinto C. Mol. Microbiol. 8:369-377(1993).</p>
ABC-3		ABC 3 transport family	<p>Members of this family include receptors that mediate transmembrane signalling. These receptors can bind to a number of factors including: amphiregulin, epidermal growth factor, gp30, heparin-binding egf, insulin, insulin-like growth factor I and II, neuregulins, transforming growth factor-alpha and, and vaccinia virus growth</p> <p>Signal transduction is mediated by catalytic activity of tyrosine kinase, such as ATP + A protein tyrosine = ADP + protein tyrosine phosphate. Typically, such signal transduction have been implicated in metabolic and developmental changes, including cell fate and differentiation. Examples include instruction of follicle cells to follow a dorsal pathway of development rather than the default ventral pathway. may also bind the spitz protein. References describing these family members and their biological activities:</p> <p>Abbot et al., J. Biol. Chem. 267:10759-10763(1992);Araki et al., J. Biol. Chem. 262:16186-16191(1987); Aroian et al., EMBO J. 13:360-366(1994); Aroian et al., Nature 348:693-699(1990); Barbetti et al., Diabetes 41:408-415(1992); Bargmann et al., Nature 319:226-230(1986); Cama et al., J. Biol. Chem. 268:8060-8069(1993); Cama et al., J. Clin. Endocrinol. Metab. 73:894-901(1991); Carrera et al., Hum. Mol. Genet. 2:1437-1441(1993); Clifford et al., Genetics 137:531-550(1994); Cocozza et al., Diabetes 41:521-526(1992); Cooke et al., Biochem. Biophys. Res. Commun. 177:1113-1120(1991); Coussens et al., Science 230:1132-1139(1985); Dickens et al., Biochem. Biophys. Res. Commun. 186:244-250(1992); Ebina et al., Cell 40:747-758(1985); Ebina et al., Proc. Natl. Acad. Sci. U.S.A. 84:704-708(1987); Ehsani et al., Genomics 15:426-429(1993); Elbein et al., Diabetes 42:429-434(1993); Elbein, Diabetes 38:737-743(1989); Fujita-Yamaguchi et al., Protein Seq. Data Anal. 1:3-6(1987); Gullick et al., EMBO J. 11:43-48(1992); Haruta et al., Diabetes 42:1837-1844(1993); Hubbard et al., EMBO J. 16:5572-5581(1997).</p> <p>Hubbard et al., Nature 372:746-754(1994); Iwanishi et al., Diabetologia 36:414-422(1993); Kadowaki et al., J. Clin. Invest. 86:254-264(1990); Kadowaki et al., Science 240:787-790(1988); Kim et al., Diabetologia 35:261-266(1992); Klinkhamer et al., EMBO J. 8:2503-2507(1989); Kusari et al., J. Biol. Chem. 266:5260-5267(1991); Lai et al., Neuron 6:691-704(1991); Lax et al., Mol. Cell. Biol. 8:1970-1978(1988); Lebrun et al., J. Biol. Chem. 268:11272-11277(1993); Lee et al., Oncogene 8:3403-3410(1993); Lesokhin et al., Dev. Biol. 205:129-144(1999); Livneh et al., Cell 40:599-607(1985).</p> <p>Longo et al., Proc. Natl. Acad. Sci. U.S.A. 90:60-64(1993); McKeon et al., Mol. Endocrinol. 4:647-656(1990); Moller et al., J. Biol. Chem. 265:14979-14985(1990); Moller et al., Mol. Endocrinol. 4:1183-1191(1990); Odawara et al., Science 245:66-68(1989); Raz et al., Genetics 129:191-201(1991).</p> <p>Sakai et al., J. Mol. Biol. 256:548-555(1996); Schaeffer et al., Biochem. Biophys. Res. Commun. 189:650-653(1992); Schejter et al., Cell 46:1091-1101(1986); Seino et al., Biochem. Biophys. Res. Commun. 159:312-316(1989); Seino et al., Diabetes 39:123-128(1990); Semba et al., Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985); Shier et al., J. Biol. Chem. 264:14605-14608(1989); Taira et al., Science 245:63-66(1989); Tewari et al., J. Biol.</p>

Pfam	Prosite	Full Name	Description
			Chem. 264:16238-16245(1989); Ullrich et al., Nature 313:756-761(1985). Ullrich et al., EMBO J. 5:2503-2512(1986); van der Vorm et al., Diabetologia 36:172-174(1993); van der Vorm et al., J. Biol. Chem. 267:66-71(1992); Wadsworth et al., Nature 314:178-180(1985); White et al., Cell 54:641-649(1988); Xu et al., J. Biol. Chem. 265:18673-18681(1990); Yamamoto et al., Nature 319:230-234(1986); and Yoshimasa et al., Science 240:784-787(1988).
ACAT		Sterol O-acyltransferase	Accession number: PF01800 Definition: Sterol O-acyltransferase Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1454 (release 4.2) Gathering cutoffs: 25 25 Trusted cutoffs: 112.80 112.80 Noise cutoffs: -128.10 -128.10 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98434592 Reference Title: Characterization of two human genes encoding acyl coenzyme Reference Title: A:cholesterol acyltransferase-related enzymes. Reference Author: Oelkers P, Behari A, Cromley D, Billheimer JT, Sturley SL; Reference Location: J Biol Chem 1998;273:26765-26771. Reference Number: [2] Reference Medline: 98434590 Reference Title: Identification of a form of acyl-CoA:cholesterol Reference Title: acyltransferase specific to liver and intestine in nonhuman Reference Title: primates. Reference Author: Anderson RA, Joyce C, Davis M, Reagan JW, Clark M, Shelness Reference Author: GS, Rudel LL; Reference Location: J Biol Chem 1998;273:26747-26754. Reference Number: [3] Reference Medline: 96243137 Reference Title: Sterol esterification in yeast: a two-gene process. Reference Author: Yang H, Bard M, Bruner DA, Gleeson A, Deckelbaum RJ, Reference Author: Aljinovic G, Pohl TM, Rothstein R, Sturley SL; Reference Location: Science 1996;272:1353-1356. Database Reference: INTERPRO; IPR002688; Comment: Sterol O-acyltransferases or acyl-coa:cholesterol acyltransferase Comment: (ACAT) EC:2.3.1.26 is a transmembrane protein that catalyses the Comment: esterification of cholesterol to its cholesterol ester storage Comment: form. Number of members: 21
ACPS		4'-phosphopantetheinyl transferase superfamily	Accession number: PF01648 Definition: 4'-phosphopantetheinyl transferase superfamily Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1679 (release 4.1) Gathering cutoffs: 0 0 Trusted cutoffs: 0.60 0.60 Noise cutoffs: -4.00 -4.00 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96027548 Reference Title: Cloning, overproduction, and

Pfam	Prosite	Full Name	Description
			<p>characterization of the</p> <p>Reference Title: Escherichia coli holo-acyl carrier protein synthase.</p> <p>Reference Author: Lambalot RH, Walsh CT;</p> <p>Reference Location: J Biol Chem 1995;270:24658-24661.</p> <p>Reference Number: [2]</p> <p>Reference Medline: 97144264</p> <p>Reference Title: A new enzyme superfamily - the phosphopantetheinyl</p> <p>Reference Title: transferases.</p> <p>Reference Author: Lambalot RH, Gehring AM, Flugel RS, Zuber P, LaCelle M,</p> <p>Reference Author: Marahiel MA, Reid R, Khosla C, Walsh CT;</p> <p>Reference Location: Chem Biol 1996;3:923-936.</p> <p>Reference Number: [3]</p> <p>Reference Medline: 10581256</p> <p>Reference Title: Crystal structure of the surfactin</p> <p>Reference Title: synthetase-activating</p> <p>Reference Title: enzyme sfp: a prototype of the 4'-phosphopantetheinyl</p> <p>Reference Title: transferase superfamily [In Process Citation]</p> <p>Reference Author: Reuter K, Mofid MR, Marahiel MA, Ficner R;</p> <p>Reference Location: EMBO J 1999;18:6823-6831.</p> <p>Database Reference: INTERPRO; IPR002582;</p> <p>Database reference: PFAMB; PB007908;</p> <p>Database reference: PFAMB; PB041384;</p> <p>Comment: Members of this family transfers the</p> <p>Comment: 4'-phosphopantetheine (4'-PP) moiety from</p> <p>Comment: coenzyme A (CoA) to</p> <p>Comment: the invariant serine of pp-binding. This post-translational</p> <p>Comment: modification renders holo-ACP capable of</p> <p>Comment: acyl group activation</p> <p>Comment: via thioesterification of the cysteamine thiol</p> <p>Comment: of 4'-PP [1].</p> <p>Comment: This superfamily consists of two subtypes:</p> <p>Comment: The ACPS type</p> <p>Comment: such as Swiss:P24224 and the Sfp type</p> <p>Comment: such as Swiss:P39135.</p> <p>Comment: The structure of the Sfp type is known [3],</p> <p>Comment: which shows the</p> <p>Comment: active site accommodates a magnesium ion.</p> <p>Comment: The most highly</p> <p>Comment: conserved regions of the alignment are</p> <p>Comment: involved in binding</p> <p>Comment: the magnesium ion.</p> <p>Number of members: 46</p>
ACT		ACT domain	<p>Accession number: PF01842</p> <p>Definition: ACT domain</p> <p>Author: Bateman A</p> <p>Alignment method of seed: Manual</p> <p>Source of seed members: Bateman A</p> <p>Gathering cutoffs: 25 0</p> <p>Trusted cutoffs: 26.10 0.50</p> <p>Noise cutoffs: 24.50 24.50</p> <p>HMM build command line: hmmbuild HMM SEED</p> <p>HMM build command line: hmmcalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 95236205</p> <p>Reference Title: The allosteric ligand site in the Vmax-type</p> <p>Reference Title: cooperative</p> <p>Reference Title: enzyme phosphoglycerate dehydrogenase.</p> <p>Reference Author: Schuller DJ, Grant GA, Banaszak LJ;</p> <p>Reference Location: Nat Struct Biol 1995;2:69-76.</p> <p>Reference Number: [2]</p> <p>Reference Medline: 99241053</p> <p>Reference Title: Gleaning non-trivial structural, functional</p> <p>Reference Title: and</p> <p>Reference Title: evolutionary information about proteins by</p> <p>Reference Title: iterative</p>

Pfam	Prosite	Full Name	Description
			<p>Reference Title: database searches.</p> <p>Reference Author: Aravind L, Koonin EV;</p> <p>Reference Location: J Mol Biol 1999;287:1023-1040.</p> <p>Database Reference: SCOP; 1psd; fa; [SCOP-USA][CATH-PDBSUM]</p> <p>Database Reference INTERPRO; IPR002912;</p> <p>Database Reference PDB; 1phz A; 35; 110;</p> <p>Database Reference PDB; 2phm A; 35; 110;</p> <p>Database Reference PDB; 1psd A; 338; 410;</p> <p>Database Reference PDB; 1psd B; 338; 410;</p> <p>Database reference: PFAMB; PB001977;</p> <p>Database reference: PFAMB; PB008097;</p> <p>Database reference: PFAMB; PB010480;</p> <p>Database reference: PFAMB; PB011031;</p> <p>Database reference: PFAMB; PB031880;</p> <p>Database reference: PFAMB; PB038464;</p> <p>Database reference: PFAMB; PB040963;</p> <p>Database reference: PFAMB; PB041518;</p> <p>Database reference: PFAMB; PB041667;</p> <p>Comment: This family of domains generally have a regulatory role.</p> <p>Comment: ACT domains are linked to a wide range of metabolic</p> <p>Comment: enzymes that are regulated by amino acid concentration.</p> <p>Comment: Pairs of ACT domains bind specifically to a particular</p> <p>Comment: amino acid leading to regulation of the linked enzyme.</p> <p>Comment: The ACT domain is found in:</p> <p>Comment: D-3-phosphoglycerate dehydrogenase EC:1.1.1.95 Swiss:P08328,</p> <p>Comment: which is inhibited by serine [1].</p> <p>Comment: Aspartokinase EC:2.7.2.4 Swiss:P53553, which is regulated by lysine.</p> <p>Comment: Acetolactate synthase small regulatory subunit Swiss:P00894,</p> <p>Comment: which is inhibited by valine.</p> <p>Comment: Phenylalanine-4-hydroxylase EC:1.14.16.1 Swiss:P00439, which</p> <p>Comment: is regulated by phenylalanine.</p> <p>Comment: Prephenate dehydrogenase EC:4.2.1.51 Swiss:P21203.</p> <p>Comment: formyltetrahydrofolate deformylase EC:3.5.1.10, Swiss:P37051,</p> <p>Comment: which is activated by methionine and inhibited by glycine.</p> <p>Comment: GTP pyrophosphokinase EC:2.7.6.5 Swiss:P11585.</p> <p>Number of members: 177</p>
Activin_rec		Activin types I and II receptor domain	<p>Accession number: PF01064</p> <p>Definition: Activin types I and II receptor domain</p> <p>Author: Finn RD, Bateman A</p> <p>Alignment method of seed: Clustalw_manual</p> <p>Source of seed members: Pfam-B_338 (release 3.0)</p> <p>Gathering cutoffs: 22 22</p> <p>Trusted cutoffs: 23.10 23.10</p> <p>Noise cutoffs: 11.30 21.20</p> <p>HMM build command line: hmmbuild -F HMM SEED</p> <p>HMM build command line: hmmcalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 97454714</p> <p>Reference Title: From receptor to nucleus: the Smad pathway.</p> <p>Reference Author: Baker JC, Harland RM;</p> <p>Reference Location: Curr Opin Genet Dev 1997;7:467-473.</p> <p>Reference Number: [2]</p> <p>Reference Medline: 94131268</p> <p>Reference Title: The TGF-beta superfamily: new members, new receptors, and</p> <p>Reference Title: new genetic tests of function in different organisms.</p> <p>Reference Author: Kingslev DM;</p>

Pfam	Prosite	Full Name	Description
			<p>Reference Location: Genes Dev 1994;8:133-146. Reference Number: [3] Reference Medline: 93390967 Reference Title: Activin receptor-like kinases: a novel subclass of Reference Title: cell-surface receptors with predicted serine/threonine Reference Title: kinase activity. Reference Author: ten Dijke P, Ichijo H, Franzen P, Schulz P, Saras J, Reference Author: Toyoshima H, Heldin CH, Miyazono K; Reference Location: Oncogene 1993;8:2879-2887. Database Reference: INTERPRO; IPR000472; Database reference: PFAMB; PB024112; Database reference: PFAMB; PB040755; Comment: This Pfam entry consists of both TGF-beta receptor types. Comment: This is an alignment of the hydrophilic cysteine-rich Comment: ligand-binding domains, Comment: Both receptor types, (type I and II) possess a 9 amino Comment: acid cysteine box, with the the consensus CCX{4-5}CN. Comment: The type I receptors also possess 7 extracellular residues Comment: preceding the cysteine box. Number of members: 79</p>
Acyl-ACP_TE		Acyl-ACP thioesterase	<p>Accession number: PF01643 Definition: Acyl-ACP thioesterase Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_928 (release 4.1) Gathering cutoffs: 25 25 Trusted cutoffs: 91.70 91.70 Noise cutoffs: -192.80 -192.80 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild --seed 0 HMM Reference Number: [1] Reference Medline: 96068671 Reference Title: Modification of the substrate specificity of an acyl-acyl Reference Title: carrier protein thioesterase by protein engineering. Reference Author: Yuan L, Voelker TA, Hawkins DJ; Reference Location: Proc Natl Acad Sci U S A 1995;92:10639-10643. Reference Number: [2] Reference Medline: 92320297 Reference Title: Fatty acid biosynthesis redirected to medium chains in Reference Title: transgenic oilseed plants. Reference Author: Voelker TA, Worrell AC, Anderson L, Bleibaum J, Fan C, Reference Author: Hawkins DJ, Radke SE, Davies HM; Reference Location: Science 1992;257:72-74. Database Reference: INTERPRO; IPR002864; Comment: This family consists of various acyl-acyl carrier protein (ACP) Comment: thioesterases (TE) these terminate fatty acyl group extension via Comment: hydrolyzing an acyl group on a fatty acid [1]. Number of members: 30</p>
Acyltransferase		Acyltransferase	<p>Accession number: PF01553 Definition: Acyltransferase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_128 (release 4.0) Gathering cutoffs: 8 8 Trusted cutoffs: 14.40 14.40 Noise cutoffs: 2.50 2.50</p>

Pfam	Prosite	Full Name	Description
			<p>HMM build command line: hmmbuild -F HMM SEED</p> <p>HMM build command line: hmmscalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 97411131</p> <p>Reference Title: Barth syndrome may be due to an acyltransferase deficiency.</p> <p>Reference Author: Neuwald AF;</p> <p>Reference Location: Curr Biol 1997;7:465-466.</p> <p>Reference Number: [2]</p> <p>Reference Medline: 96224398</p> <p>Reference Title: A novel X-linked gene, G4.5, is responsible for Barth</p> <p>Reference Title: syndrome.</p> <p>Reference Author: Bione S, D'Adamo P, Maestrini E, Gedeon AK, Bolhuis PA,</p> <p>Reference Author: Toniolo D;</p> <p>Reference Location: Nat Genet 1996;12:385-389.</p> <p>Database Reference INTERPRO; IPR002123;</p> <p>Database reference: PFAMB; PB009622;</p> <p>Database reference: PFAMB; PB009717;</p> <p>Database reference: PFAMB; PB033259;</p> <p>Database reference: PFAMB; PB041102;</p> <p>Database reference: PFAMB; PB041638;</p> <p>Comment: This family contains acyltransferases involved in phospholipid</p> <p>Comment: biosynthesis and other proteins of unknown function [1]. This</p> <p>Comment: family also includes tafazzin Swiss:Q16635, the Barth syndrome</p> <p>Comment: gene [2].</p> <p>Number of members: 74</p>
Adaptin_N		Adaptin N terminal region	<p>Accession number: PF01602</p> <p>Definition: Adaptin N terminal region</p> <p>Author: Bashton M, Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_491 (release 4.0)</p> <p>Gathering cutoffs: 12 12</p> <p>Trusted cutoffs: 15.50 15.50</p> <p>Noise cutoffs: 9.00 9.00</p> <p>HMM build command line: hmmbuild -f HMM SEED</p> <p>HMM build command line: hmmscalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 97409270</p> <p>Reference Title: Linking cargo to vesicle formation: receptor tail</p> <p>Reference Title: interactions with coat proteins.</p> <p>Reference Author: Kirchhausen T, Bonifacino JS, Riezman H;</p> <p>Reference Location: Curr Opin Cell Biol 1997;9:488-495.</p> <p>Reference Number: [2]</p> <p>Reference Medline: 89202379</p> <p>Reference Title: Structural and functional division into two domains of the</p> <p>Reference Title: large (100- to 115-kDa) chains of the clathrin-associated</p> <p>Reference Title: protein complex AP-2.</p> <p>Reference Author: RAKirchhausen T, Nathanson KL, Matsui W, Vaisberg A, Chow</p> <p>Reference Author: EP, Burne C, Keen JH, Davis AE;</p> <p>Reference Location: Proc Natl Acad Sci U S A 1989;86:2612-2616.</p> <p>Database Reference INTERPRO; IPR002553;</p> <p>Database reference: PFAMB; PB040953;</p> <p>Comment: This family consists of the N terminal region of various alpha,</p> <p>Comment: beta and gamma subunits of the AP-1, AP-2 and AP-3 adaptor</p> <p>Comment: protein complexes. The adaptor protein (AP) complexes are involved in</p> <p>Comment: the formation of clathrin-coated pits and vesicles [1].</p> <p>Comment: The N-terminal region of the various adaptor proteins (APs) is constant</p>

0966980 101300

Pfam	Prosite	Full Name	Description
			<p>Comment: by comparison to the C-terminal which is variable within members of the</p> <p>Comment: AP-2 family[2]; and it has been proposed that this constant region</p> <p>Comment: interacts with another uniform component of the coated vesicles [2].</p> <p>Number of members: 66</p>
ALAD	PDOC00153	Delta-aminolevulinic acid dehydratase active site	<p>Delta-aminolevulinic acid dehydratase (EC 4.2.1.24) (ALAD) [1] catalyzes the second step in the biosynthesis of heme, the condensation of two molecules of 5-aminolevulinate to form porphobilinogen. The enzyme is an oligomer composed of eight identical subunits. Each of the subunits binds an atom of zinc or of magnesium (in plants). A lysine has been implicated in the catalytic mechanism [2]. The sequence of the region in the vicinity of the active site residue is conserved in ALAD from various prokaryotic and eukaryotic species.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern G-x-D-x-[LIVM](2)-[IV]-K-P-[GSA]-x(2)-Y [K is the active site residue]</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update November 1995 / Pattern and text revised.</p> <p>References</p> <p>[1] Li J.-M., Russell C.S., Cosloy S.D. Gene 75:177-184(1989).</p> <p>[2] Gibbs P.N.B., Jordan P.M. Biochem. J. 236:447-451(1986).</p>
Aldolase	PDOC00144	KDPG and KHG aldolases active site signatures	<p>4-hydroxy-2-oxoglutarate aldolase (EC 4.1.3.16) (KHG-aldolase) catalyzes the interconversion of 4-hydroxy-2-oxoglutarate into pyruvate and glyoxylate.</p> <p>Phospho-2-dehydro-3-deoxygluconate aldolase (EC 4.1.2.14) (KDPG-aldolase) catalyzes the interconversion of 6-phospho-2-dehydro-3-deoxy-D-gluconate into pyruvate and glyceraldehyde 3-phosphate.</p> <p>These two enzymes are structurally and functionally related [1]. They are both homotrimeric proteins of approximately 220 amino-acid residues. They are class I aldolases whose catalytic mechanism involves the formation of a Schiff-base intermediate between the substrate and the epsilon-amino group of a lysine residue. In both enzymes, an arginine is required for catalytic activity.</p> <p>We developed two signature patterns for these enzymes. The first one contains the active site arginine and the second, the lysine involved in the Schiff-base formation.</p> <p>Description of pattern(s) and/or profile(s)</p>

Pfam	Prosite	Full Name	Description
			<p>Consensus pattern G-[LIVM]-x(3)-E-[LIV]-T-[LF]-R [R is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for <i>Bacillus subtilis</i> KDPG-aldolase which has Thr instead of Arg in the active site. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Consensus pattern G-x(3)-[LIVMF]-K-[LF]-F-P-[SA]-x(3)-G [K is involved in Schiff-base formation] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Patterns and text revised. References [1] Vlahos C J., Dekker E.E. J. Biol. Chem. 263:11683-11691(1988).</p>
Alpha_L_fucos	PDOC00324	Alpha-L-fucosidase	<p>Alpha-L-fucosidase (EC 3.2.1.51) [1] is a lysosomal enzyme responsible for hydrolyzing the alpha-1,6-linked fucose joined to the reducing-end N-acetylglucosamine of the carbohydrate moieties of glycoproteins. Deficiency of alpha-L-fucosidase results in the lysosomal storage disease fucosidosis.</p> <p>A cysteine residue is important for the activity of the enzyme. There is only one cysteine conserved between the sequence of mammalian alpha-L-fucosidase and that of the slime mold <i>Dictyostelium discoideum</i>. We have derived a pattern from the region around that conserved cysteine.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern P-x(2)-L-x(3)-K-W-E-x-C [C is the putative active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Note these proteins belong to family 29 in the classification of glycosyl hydrolases [2,E1]. Last update November 1997 / Pattern and text revised. References [1] Fisher K.J., Aronson N.N. Jr. Biochem. J. 264:695-701(1989).</p> <p>[2] Henrissat B. Biochem. J. 280:309-316(1991).</p> <p>[E1] http://www.expasy.ch/cgi-bin/lists?glycosid.txt</p>
Amino_oxidase		Flavin containing amine oxidase	<p>Accession number: PF01593 Definition: Flavin containing amine oxidase Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_606 (release 4.1) Gathering cutoffs: -110 -110 Trusted cutoffs: -110.00 -110.00 Noise cutoffs: -111.80 -111.80 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98258926</p>

Pfam	Prosite	Full Name	Description
			<p>Reference Title: Maize polyamine oxidase: primary structure from protein and</p> <p>Reference Title: cDNA sequencing.</p> <p>Reference Author: Tavladoraki P, Schinina ME, Cecconi F, Agostino SD, Manera</p> <p>Reference Author: F, Rea G, Mariottini P, Federico R, Angelini R;</p> <p>Reference Location: FEBS Lett 1998;426:62-66.</p> <p>Reference Number: [2]</p> <p>Reference Medline: 97306298</p> <p>Reference Title: A key amino acid responsible for substrate selectivity of</p> <p>Reference Title: monoamine oxidase A and B.</p> <p>Reference Author: Tsugeno Y, Ito A;</p> <p>Reference Location: J Biol Chem 1997;272:14033-14036.</p> <p>Reference Number: [3]</p> <p>Reference Medline: 95287865</p> <p>Reference Title: Cloning, sequencing and heterologous expression of the</p> <p>Reference Title: monoamine oxidase gene from <i>Aspergillus niger</i>.</p> <p>Reference Author: Schilling B, Lerch K;</p> <p>Reference Location: Mol Gen Genet 1995;247:430-438.</p> <p>Database Reference: SCOP; 1b37; fa; [SCOP-USA][CATH-PDBSUM]</p> <p>Database Reference: INTERPRO; IPR002937;</p> <p>Database Reference: PDB; 1b37 A; 14; 455;</p> <p>Database Reference: PDB; 1b5q A; 14; 455;</p> <p>Database Reference: PDB; 1b37 B; 14; 455;</p> <p>Database Reference: PDB; 1b37 C; 14; 455;</p> <p>Database Reference: PDB; 1b5q B; 14; 455;</p> <p>Database Reference: PDB; 1b5q C; 14; 455;</p> <p>Database reference: PFAMB; PB017518;</p> <p>Database reference: PFAMB; PB024839;</p> <p>Database reference: PFAMB; PB040747;</p> <p>Comment: This family consists of various amine oxidases, including maize polyamine</p> <p>Comment: oxidase (PAO) [1] and various flavin containing monoamine oxidases</p> <p>Comment: (MAO). The aligned region includes the flavin binding site of these</p> <p>Comment: enzymes.</p> <p>Comment: In vertebrates MAO plays an important role regulating the intracellular</p> <p>Comment: levels of amines via their oxidation; these include various</p> <p>Comment: neurotransmitters, neurotoxins and trace amines [2]. In lower eukaryotes</p> <p>Comment: such as <i>Aspergillus</i> and in bacteria the main role of amine oxidases is</p> <p>Comment: to provide a source of ammonium [3].</p> <p>Comment: PAOs in plants, bacteria and protozoa oxidase spermidine and spermine</p> <p>Comment: to an aminobutylal, diaminopropane and hydrogen peroxide and are</p> <p>Comment: involved in the catabolism of polyamines [1].</p> <p>Comment: Other members of this family include tryptophan 2-monooxygenase,</p> <p>Comment: putrescine oxidase, corticosteroid binding proteins and antibacterial</p> <p>Comment: glycoproteins.</p> <p>Number of members: 58</p>
ANF_receptor	PDOC00430	Natriuretic peptides receptors signature	<p>Natriuretic peptides are hormones involved in the regulation of fluid and electrolyte homeostasis. These hormones stimulate the intracellular production of cyclic GMP as a second messenger.</p> <p>Currently, three types of natriuretic peptide receptors are known [1,2]. Two express guanylate cyclase activity: GC-A (or ANP-A) which seems specific to atrial natriuretic peptide (ANP), and GC-B (or ANP-B) which</p>

Pfam	Prosite	Full Name	Description
			<p>seems to be stimulated more effectively by brain natriuretic peptide (BNP) than by ANP. The third receptor (ANP-C) is probably responsible for the clearance of ANP from the circulation and does not play a role in signal transduction.</p> <p>GC-A and GC-B are plasma membrane-bound proteins that share the following topology: an N-terminal extracellular domain which acts as the ligand binding region, then a transmembrane domain followed by a large cytoplasmic C-terminal region that can be subdivided into two domains: a protein kinase-like domain (see <PDOC00100>) that appears important for proper signalling and a guanylate cyclase catalytic domain (see <PDOC00425>). The topology of ANP-C is different: like GC-A and -B it possesses an extracellular ligand-binding region and a transmembrane domain, but its cytoplasmic domain is very short.</p> <p>We developed a pattern from the ligand-binding region of natriuretic peptide receptors based on a highly conserved region located in the N-terminal part of the domain.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern G-P-x-C-x-Y-x-A-A-x-V-x-R-x(3)-H-W Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update May 1991 / First entry. References [1] Garbers D.L. New Biol. 2:499-504(1990).</p> <p>[2] Schulz S., Chinkers M., Garbers D.L. FASEB J. 2:2026-2035(1989).</p>
Apocytochrome_F	PDOC00169	Cytochrome c family heme-binding site signature	<p>In proteins belonging to cytochrome c family [1], the heme group is covalently attached by thioether bonds to two conserved cysteine residues. The consensus sequence for this site is Cys-X-X-Cys-His and the histidine residue is one of the two axial ligands of the heme iron. This arrangement is shared by all proteins known to belong to cytochrome c family, which presently includes cytochromes c, c', c1 to c6, c550 to c556, cc3/Hmc, cytochrome f and reaction center cytochrome c.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern C-{CPWHF}-{CPWR}-C-H-{CFYW} Sequences known to belong to this class detected by the pattern ALL, except for four cytochrome c's which lack the first thioether bond. Other sequence(s) detected in SWISS-PROT 454.</p>

Pfam	Prosite	Full Name	Description
			<p>Note: some cytochrome c's have more than a single bound heme group c4 has 2, c7 has 3, c3 has 4, the reaction center has 4, and cc3/Hmc has 16 !</p> <p>Last update June 1992 / Text revised.</p> <p>References [1] Mathews F.S. Prog. Biophys. Mol. Biol. 45:1-56(1985).</p>
arf	PDOC00781 PDOC00017 PDOC01020	ADP-ribosylation factors family signature; ATP/GTP-binding site motif A (P-loop); ATP phosphoribosyltransferase signature PROSITE cross-reference(s)	<p>ADP-ribosylation factors (ARF) [1,2,3,4] are 20 Kd GTP-binding proteins involved in protein trafficking. They may modulate vesicle budding and uncoating within the Golgi apparatus. ARF's also act as allosteric activators of cholera toxin ADP-ribosyltransferase activity. They are evolutionary conserved and present in all eukaryotes. At least six forms of ARF are present in mammals and three in budding yeast. The ARF family also includes proteins highly related to ARF's but which lack the cholera toxin cofactor activity, they are collectively known as ARL's (ARF-like).</p> <p>ARD1 is a 64 Kd mammalian protein of unknown biological function that contains an ARF domain at its C-terminal extremity.</p> <p>Proteins from the ARF family are generally included in the RAS 'superfamily' of small GTP-binding proteins [5], but they are only slightly related to the other RAS proteins. They also differ from RAS proteins in that they lack cysteine residues at their C-termini and are therefore not subject to prenylation. The ARFs are N-terminally myristoylated (the ARLs have not yet been shown to be modified in such a fashion).</p> <p>As a signature pattern, we selected a conserved region in the C-terminal part of ARF's and ARL's.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [HRQT]-x-[FYWI]-x-[LIVM]-x(4)-A-x(2)-G-x(2)-[LIVM]-x(2)-[GSA]-[LIVMF]-x-[WK]-[LIVM] Sequences known to belong to this class detected by the pattern ALL, except for 4 sequences. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Note proteins belonging to this family also contain a copy of the ATP/GTP-binding motif 'A' (P-loop) (see <PDOC00017>). Expert(s) to contact by email Kahn R.A. rkahn@bimcore.emory.edu</p> <p>Last update November 1997 / Pattern and text revised. Cell. Signal. 4:367-399(1993). References [1] Boman A.L., Kahn R.A. Trends Biochem. Sci. 20:147-150(1995).</p> <p>[2] Moss J., Vaughan M.</p> <p>[3] Moss J., Vaughan M.</p>

Pfam	Prosite	Full Name	Description
			<p>pattern: in the last position Gly is found instead of Ser or Thr.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [AG]-x(4)-G-K-[ST] Sequences known to belong to this class detected by the pattern a majority. Other sequence(s) detected in SWISS-PROT in addition to the proteins listed above, the 'A' motif is also found in a number of other proteins. Most of these proteins probably bind a nucleotide, but others are definitively not ATP- or GTP-binding (as for example chymotrypsin, or human ferritin light chain). Expert(s) to contact by email Koonin E.V. koonin@ncbi.nlm.nih.gov</p> <p>Last update July 1999 / Text revised.</p> <p>References</p> <p>[1] Walker J.E., Saraste M., Runswick M.J., Gay N.J. EMBO J. 1:945-951(1982).</p> <p>[2] Moller W., Amons R. FEBS Lett. 186:1-7(1985).</p> <p>[3] Fry D.C., Kuby S.A., Mildvan A.S. Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986).</p> <p>[4] Dever T.E., Glynnias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987).</p> <p>[5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Biochem. Sci. 15:430-434(1990).</p> <p>[6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993).</p> <p>[7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990).</p> <p>[8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata).</p> <p>[9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989).</p> <p>[10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989).</p> <p>ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we selected a region located in the C-terminal part of this enzyme.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-G-x-T-[LM]</p>

Pfam	Prosite	Full Name	Description
			<p>Comment: these enzymes catalyse hydrolytic dechlorination of their substrates.</p> <p>Comment: Atrazine chlorohydrolase (AtzA) from <i>Pseudomonas</i> sp. Swiss:P72156</p> <p>Comment: catalyses the dechlorination of atrazine to hydroxyatrazine [1].</p> <p>Comment: s-Triazine hydrolase (TrzA) from <i>R. corallinus</i> Swiss:P72156</p> <p>Comment: catalyses the deamination and dechlorination of melamine and</p> <p>Comment: deethylsimazine to ammeline and N-ethylammeline [1].</p> <p>Number of members: 29</p>
B56		Protein phosphatase 2A regulatory B subunit (B56 family)	<p>Accession number: PF01603</p> <p>Definition: Protein phosphatase 2A regulatory B subunit (B56 family)</p> <p>Author: Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_984 (release 4.1)</p> <p>Gathering cutoffs: 11 11</p> <p>Trusted cutoffs: 17.80 17.80</p> <p>Noise cutoffs: 5.50 5.50</p> <p>HMM build command line: hmmbuild -f HMM SEED</p> <p>HMM build command line: hmmscalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 96064678</p> <p>Reference Title: Identification of a new family of protein phosphatase 2A regulatory subunits.</p> <p>Reference Author: McCright B, Virshup DM;</p> <p>Reference Location: J Biol Chem 1995;270:26123-26128.</p> <p>Database Reference: INTERPRO; IPR002554;</p> <p>Comment: Protein phosphatase 2A (PP2A) is a major intracellular protein</p> <p>Comment: phosphatase that regulates multiple aspects of cell growth and metabolism.</p> <p>Comment: The ability of this widely distributed heterotrimeric enzyme to act on a</p> <p>Comment: diverse array of substrates is largely controlled by the nature of its</p> <p>Comment: regulatory B subunit. There are multiple families of B subunits (See also</p> <p>Comment: PR55), this family is called the B56 family [1].</p> <p>Number of members: 34</p>
Bac_export_1		Bacterial export proteins, family 1	<p>Accession number: PF01311</p> <p>Definition: Bacterial export proteins, family 1</p> <p>Author: Finn RD, Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_1442 (release 3.0)</p> <p>Gathering cutoffs: 25 25</p> <p>Trusted cutoffs: 37.20 37.20</p> <p>Noise cutoffs: -95.00 -95.00</p> <p>HMM build command line: hmmbuild -F HMM SEED</p> <p>HMM build command line: hmmscalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 95113771</p> <p>Reference Title: Caulobacter FliQ and FliR membrane proteins, required for</p> <p>Reference Title: flagellar biogenesis and cell division, belong to a family</p> <p>Reference Title: of virulence factor export proteins.</p> <p>Reference Author: Zhuang WY, Shapiro L;</p> <p>Reference Location: J Bacteriol 1995;177:343-356.</p> <p>Database Reference: INTERPRO; IPR002010;</p> <p>Comment: This family includes the following members;</p> <p>Comment: FliR, MopE, SsaT, YopT, Hrp, HrcT and SpaR</p> <p>Comment: All of these members export proteins, that do not possess signal</p> <p>Comment: peptides, through the membrane. Although the proteins that these</p>

0066980 10300

Pfam	Prosite	Full Name	Description
			<p>conserved positions found at the N-terminal extremity of the domain, the second is located in the C-terminal section.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern W-[LIV]-x(3)-[KRQ]-x-[LIVM]-x(2)-[QH]-x(0,2)-[LIVMF]-x(6,8)-[LIVMF]-x(3,5)-F-[FY]-x(2)-[DENS] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Consensus pattern [HYW]-x(9)-[DENQSTV]-[SA]-x(3)-[FY]-[LIVM]-x(2)-[ACV]-x(2)-[LM]-x(2)-[FY]-G-x-[DENQST]-[LIVMFYS] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT 7.</p> <p>Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Expert(s) to contact by email Rees J. jrees@vax.oxford.ac.uk</p> <p>Last update November 1997 / Patterns and text revised; profile added. References [1] Rees D.J.G., Ades S.A., Singer S.J., Hynes R.O. Nature 347:685-689(1990). [2] Funayama N., Nagafuchi A., Sato N., Tsukita S., Tsukita S. J. Cell Biol. 115:1039-1048(1991). [3] Takeuchi K., Kawashima A., Nagafuchi A., Tsukita S. J. Cell Sci. 107:1921-1928(1994).</p>
biotin_lipoyl	PDOC00167; PDOC00168	Biotin-requiring enzymes; 2-oxo acid dehydrogenases acyltransferase component lipoyl binding	<p>Biotin, which plays a catalytic role in some carboxyl transfer reactions, is covalently attached, via an amide bond, to a lysine residue in enzymes requiring this coenzyme [1,2,3,4]. Such enzymes are:</p> <ul style="list-style-type: none"> - Pyruvate carboxylase (EC 6.4.1.1). - Acetyl-CoA carboxylase (EC 6.4.1.2). - Propionyl-CoA carboxylase (EC 6.4.1.3). - Methylcrotonoyl-CoA carboxylase (EC 6.4.1.4). - Geranoyl-CoA carboxylase (EC 6.4.1.5). - Urea carboxylase (EC 6.3.4.6). - Oxaloacetate decarboxylase (EC 4.1.1.3). - Methylmalonyl-CoA decarboxylase (EC 4.1.1.41). - Glutaconyl-CoA decarboxylase (EC 4.1.1.70). - Methylmalonyl-CoA carboxyl-transferase (EC 2.1.3.1) (transcarboxylase). <p>Sequence data reveal that the region around the biocytin (biotin-lysine) residue is well conserved and can be used as a signature pattern.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [GN]-[DEQTR]-x-[LIVMFY]-x(2)-[LIVM]-x-[AIV]-M-K-[LMAT]-x(3)-[LIVM]-x-[SAV] [K is the biotin attachment site]</p>

Pfam	Prosite	Full Name	Description
			<p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Note the domain around the biotin-binding lysine residue is evolutionary related to that around the lipoyl-binding lysine residue of 2-oxo acid dehydrogenase acyltransferases (see <PDOC00168>).</p> <p>Last update November 1997 / Pattern and text revised.</p> <p>References [1] Knowles J.R. Annu. Rev. Biochem. 58:195-221(1989).</p> <p>[2] Samols D., Thronton C.G., Murtif V.L., Kumar G.K., Haase F.C., Wood H.G. J. Biol. Chem. 263:6461-6464(1988).</p> <p>[3] Goss N.H., Wood H.G. Meth. Enzymol. 107:261-278(1984).</p> <p>[4] Shenoy B.C., Xie Y., Park V.L., Kumar G.K., Beegen H., Wood H.G., Samols D. J. Biol. Chem. 267:18407-18412(1992).</p> <p>The 2-oxo acid dehydrogenase multienzyme complexes [1,2] from bacterial and eukaryotic sources catalyze the oxidative decarboxylation of 2-oxo acids to the corresponding acyl-CoA. The three members of this family of multienzyme complexes are:</p> <ul style="list-style-type: none"> - Pyruvate dehydrogenase complex (PDC). - 2-oxoglutarate dehydrogenase complex (OGDC). - Branched-chain 2-oxo acid dehydrogenase complex (BCOADC). <p>These three complexes share a common architecture: they are composed of multiple copies of three component enzymes - E1, E2 and E3. E1 is a thiamine pyrophosphate-dependent 2-oxo acid dehydrogenase, E2 a dihydrolipamide acyltransferase, and E3 an FAD-containing dihydrolipamide dehydrogenase.</p> <p>E2 acyltransferases have an essential cofactor, lipoic acid, which is covalently bound via an amide linkage to a lysine group. The E2 components of OGDC and BCOADC bind a single lipoyl group, while those of PDC bind either one (in yeast and in Bacillus), two (in mammals), or three (in Escherichia coli) lipoyl groups [3].</p> <p>In addition to the E2 components of the three enzymatic complexes described above, a lipoic acid cofactor is also found in the following proteins:</p> <ul style="list-style-type: none"> - H-protein of the glycine cleavage system (GCS) [4]. GCS is a multienzyme complex of four protein components, which catalyzes the degradation of glycine. H protein shuttles the methylamine group of glycine from the P protein to the T protein. H-protein from either prokaryotes or eukaryotes binds a single lipoic group.

Pfam	Prosite	Full Name	Description
			<p>- Mammalian and yeast pyruvate dehydrogenase complexes differ from that of other sources, in that they contain, in small amounts, a protein of unknown function - designated protein X or component X. Its sequence is closely related to that of E2 subunits and seems to bind a lipoic group [5].</p> <p>- Fast migrating protein (FMP) (gene acoC) from <i>Alcaligenes eutrophus</i> [6].</p> <p>This protein is most probably a dihydrolipamide acyltransferase involved in acetoin metabolism.</p> <p>We developed a signature pattern which allows the detection of the lipoyl-binding site.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [GN]-x(2)-[LIVF]-x(5)-[LIVFC]-x(2)-[LIVFA]-x(3)-K-[STAIV]-[STAVQDN]-x(2)-[LIVMFS]-x(5)-[GCN]-x-[LIVMFY] [K is the lipoyl-binding site]</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT 2.</p> <p>Note the domain around the lipoyl-binding lysine residue is evolutionary related to that around the biotin-binding lysine residue of biotin requiring enzymes (see <PDOC00167>).</p> <p>Last update November 1995 / Text revised.</p> <p>References</p> <p>[1] Yeaman S.J. Biochem. J. 257:625-632(1989).</p> <p>[2] Yeaman S.J. Trends Biochem. Sci. 11:293-296(1986).</p> <p>[3] Russel G.C., Guest J.R. Biochim. Biophys. Acta 1076:225-232(1991).</p> <p>[4] Fujiwara K., Okamura-Ikeda K., Motokawa Y. J. Biol. Chem. 261:8836-8841(1986).</p> <p>[5] Behal R.H., Browning K.S., Hall T.B., Reed L.J. Proc. Natl. Acad. Sci. U.S.A. 86:8732-8736(1989).</p> <p>[6] Priefert H., Hein S., Krueger N., Zeh K., Schmidt B., Steinbuechel A. J. Bacteriol. 173:4056-4071(1991).</p>
Biotin_synth		Biotin synthase	<p>Accession number: PF01792</p> <p>Definition: Biotin synthase</p> <p>Author: Bashton M, Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_1407 (release 4.2)</p> <p>Gathering cutoffs: -180 -180</p> <p>Trusted cutoffs: -176.30 -176.30</p> <p>Noise cutoffs: -183.90 -183.90</p> <p>HMM build command line: hmmbuild -F HMM SEED</p> <p>HMM build command line: hmmcalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 96312354</p> <p>Reference Title: Cloning, sequencing, and characterization of the <i>Bacillus</i></p>

Pfam	Prosite	Full Name	Description
			<p>Reference Title: subtilis biotin biosynthetic operon. Reference Author: Bower S, Perkins JB, Yocum RR, Howitt CL, Rahaim P, Pero J; Reference Location: J Bacteriol 1996;178:4122-4130. Reference Number: [2] Reference Medline: 97074643 Reference Title: Two new members of the bio B superfamily: cloning, Reference Title: sequencing and expression of bio B genes of <i>Methylobacillus</i> Reference Title: flagellatum and <i>Corynebacterium</i> glutamicum. Reference Author: Serebriiskii IG, Vassin VM, Tsygankov YD; Reference Location: Gene 1996;175:15-22. Database Reference: INTERPRO; IPR002684; Database reference: PFAMB; PB023954; Database reference: PFAMB; PB040740; Database reference: PFAMB; PB041208; Comment: Biotin synthase EC:2.8.1.6 works with flavodoxin, S-adenosylmethionine, Comment: and possibly cysteine to convert dethiobiotin to biotin [1]. Comment: Biotin (vitamin H) is a prosthetic group in enzymes catalysing Comment: carboxylation and transcarboxylation reactions [2]. Number of members: 29</p>
BolA		BolA-like protein	<p>Accession number: PF01722 Definition: BolA-like protein Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1996 (release 4.1) Gathering cutoffs: 23 23 Trusted cutoffs: 23.70 23.70 Noise cutoffs: -16.00 -16.00 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 99291046 Reference Title: The stationary-phase morphogene bolA from <i>Escherichia coli</i> Reference Title: is induced by stress during early stages of growth. Reference Author: Santos JM, Freire P, Vicente M, Arraiano CM; Reference Location: Mol Microbiol 1999;32:789-798. Reference Number: [2] Reference Medline: 90059998 Reference Title: Induction of a growth-phase-dependent promoter triggers Reference Title: transcription of bolA, an <i>Escherichia coli</i> morphogene. Reference Author: Aldea M, Garrido T, Hernandez-Chico C, Vicente M, Kushner Reference Author: SR; Reference Location: EMBO J 1989;8:3923-3931. Database Reference: INTERPRO; IPR002634; Comment: This family consist of the morpho-protein BolA from Comment: <i>E. coli</i> and its various homologs. In <i>E. coli</i> over expression of Comment: this protein causes round morphology and may be involved in Comment: switching the cell between elongation and septation systems during Comment: cell division [1]. The expression of BolA is growth rate regulated Comment: and is induced during the transition into the the stationary Comment: phase [1]. BolA is also induced by stress during early stages of Comment: growth [1] and may have a general role in</p>

Pfam	Prosite	Full Name	Description
			<p>stress response.</p> <p>Comment: It has also been suggested that BolA can induce the transcription</p> <p>Comment: of penicillin binding proteins 6 and 5 [2,1].</p> <p>Number of members: 18</p>
casein_kappa			<p>Accession number: PF00997</p> <p>Definition: Kappa casein</p> <p>Author: Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_1298 (release 3.0)</p> <p>Gathering cutoffs: -32 -32</p> <p>Trusted cutoffs: 16.40 16.40</p> <p>Noise cutoffs: -73.00 -73.00</p> <p>HMM build command line: hmmbuild -F HMM SEED</p> <p>HMM build command line: hmmscalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 98072500</p> <p>Reference Title: Nucleotide sequence evolution at the kappa-casein locus:</p> <p>Reference Title: evidence for positive selection within the family Bovidae.</p> <p>Reference Author: Ward TJ, Honeycutt RL, Derr JN;</p> <p>Reference Location: Genetics 1997;147:1863-1872.</p> <p>Database Reference: INTERPRO; IPR000117;</p> <p>Comment: Kappa-casein is a mammalian milk protein involved in a</p> <p>Comment: number of important physiological processes. In the gut,</p> <p>Comment: the ingested protein is split into an insoluble peptide</p> <p>Comment: (para kappa-casein) and a soluble hydrophilic glycopeptide</p> <p>Comment: (caseinomacropeptide).</p> <p>Caseinomacropeptide is responsible</p> <p>Comment: for increased efficiency of digestion, prevention of neonate</p> <p>Comment: hypersensitivity to ingested proteins, and inhibition of</p> <p>Comment: gastric pathogens.</p> <p>Number of members: 56</p>
CAT	PDOC00093	Chloramphenicol acetyltransferase	<p>Chloramphenicol acetyltransferase (CAT) (EC 2.3.1.28) [1] catalyzes the acetyl-CoA dependent acetylation of chloramphenicol (Cm), an antibiotic which inhibits prokaryotic peptidyltransferase activity. Acetylation of Cm by CAT inactivates the antibiotic. A histidine residue, located in the C-terminal section of the enzyme, plays a central role in its catalytic mechanism. We derived a signature pattern from the region surrounding this active site residue.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern Q-[LIV]-H-H-[SA]-x(2)-D-G-[FY]-H [The second H is the active site residue]</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Note there is a second family of CAT [2], evolutionary unrelated to the main family described above. These CAT belong to the bacterial hexapeptide-repeat containing-transferases family (see <PDOC00094>).</p> <p>Last update November 1997 / Text revised.</p> <p>References [1]</p>

Pfam	Prosite	Full Name	Description
			<p>Shaw W.V., Leslie A.G.W. Annu. Rev. Biophys. Chem. 20:363-386(1991).</p> <p>[2] Parent R., Roy P.H. J. Bacteriol. 174:2891-2897(1992).</p>
Cation_efflux		Cation efflux family	<p>Accession number: PF01545 Definition: Cation efflux family Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_232 (release 4.0) Gathering cutoffs: -6 -6 Trusted cutoffs: 6.90 6.90 Noise cutoffs: -19.30 -19.30 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98361887 Reference Title: Molecular characterization of a chromosomal determinant Reference Title: conferring resistance to zinc and cobalt ions in Reference Title: Staphylococcus aureus. Reference Author: Xiong A, Jayaswal RK; Reference Location: J Bacteriol 1998;180:4024-4029. Reference Number: [2] Reference Medline: 96219090 Reference Title: Cloning and sequence analysis of czc genes in Alcaligenes Reference Title: sp. strain CT14. Reference Author: Kunito T, Kusano T, Oyaizu H, Senoo K, Kanazawa S, Reference Author: Matsumoto S; Reference Location: Biosci Biotechnol Biochem 1996;60:699-704. Database Reference: INTERPRO; IPR002524; Database reference: PFAMB; PB038216; Comment: Members of this family are integral membrane proteins, that Comment: are found to increase tolerance to divalent metal ions such Comment: as cadmium, zinc, and cobalt. These proteins are thought to Comment: be efflux pumps that remove these ions from cells. Number of members: 59</p>
CBD_6		Cellulose binding domain	<p>Accession number: PF02018 Definition: Cellulose binding domain Author: Bateman A Alignment method of seed: Manual Source of seed members: Chris Ponting Gathering cutoffs: 19 0 Trusted cutoffs: 19.10 19.10 Noise cutoffs: 8.90 8.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97074498 Reference Title: Structure of the N-terminal cellulose-binding domain of Reference Title: Cellulomonas fimi CenC determined by nuclear magnetic Reference Title: resonance spectroscopy. Reference Author: Johnson PE, Joshi MD, Tomme P, Kilburn DG, McIntosh LP; Reference Location: Biochemistry 1996;35:14381-14394. Database Reference: URL; http://www.ocms.ox.ac.uk/~ponting/methmb/example.html; Database Reference: SCOP; 1ulp; fa; [SCOP-USA][CATH-PDBSUM] Database Reference: PDB; 1ulo ; 1; 149; Database Reference: PDB; 1ulp ; 1; 149;</p>

Pfam	Prosite	Full Name	Description
			<p>Database Reference PDB; 1cx1 A; 2; 6; Database Reference PDB; 1ulo ; 150; 152; Database Reference PDB; 1ulp ; 150; 152; Database Reference PDB; 1cx1 A; 7; 151; Database reference: PFAMB; PB012497; Database reference: PFAMB; PB041237; Database reference: PFAMB; PB041605; Number of members: 76</p>
CBFD_NFYB_HMF	PDOC00578	CBF/NF-Y subunits signatures	<p>Diverse DNA binding proteins are known to bind the CCAAT box, a common cis-acting element found in the promoter and enhancer regions of a large number of genes in eukaryotes. Amongst these proteins is one known as the CCAAT-binding factor (CBF) or NF-Y [1]. CBF is a heteromeric transcription factor that consists of two different components both needed for DNA-binding.</p> <p>The HAP protein complex of yeast binds to the upstream activation site of cytochrome C iso-1 gene (CYC1) as well as other genes involved in mitochondrial electron transport and activates their expression. It also recognizes the sequence CCAAT and is structurally and evolutionary related to CBF.</p> <p>The first subunit of CBF, known as CBF-A or NF-YB in vertebrates, HAP3 in budding yeast and as php3 in fission yeast, is a protein of 116 to 210 amino-acid residues which contains a highly conserved central domain of about 90 residues. This domain seems to be involved in DNA-binding; we have developed a signature pattern from its central part.</p> <p>The second subunit of CBF, known as CBF-B or NF-YA in vertebrates, HAP2 in budding yeast and php2 in fission yeast, is a protein of 265 to 350 amino-acid residues which contains a highly conserved region of about 60 residues. This region, called the 'essential core' [2], seems to consist of two subdomains: an N-terminal subunit-association domain and a C-terminal DNA recognition domain. We have developed a signature pattern from a section of the subunit-association domain.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern C-V-S-E-x-I-S-F-[LIVM]-T-[SG]-E-A-[SC]-[DE]-[KRQ]-C Sequences known to belong to this class detected by the pattern ALL CBF-A subunits. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Consensus pattern Y-V-N-A-K-Q-Y-x-R-I-L-K-R-R-x-A-R-A-K-L-E Sequences known to belong to this class detected by the pattern ALL CBF-B subunits. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update November 1995 / Patterns and text revised. References [1] Li X.-Y., Mantovani R., Hooft van Huijsduijnen R., Andre I., Benoist C., Mathis D.</p>

00669900 08668960

Pfam	Prosite	Full Name	Description
			Nucleic Acids Res. 20:1087-1091(1992). [2] Olesen J.T., Fikes J.D., Guarente L. Mol. Cell. Biol. 11:611-619(1991).
CbiX		CbiX	Accession number: PF01903 Definition: CbiX Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: -25 -25 Trusted cutoffs: -23.10 -23.10 Noise cutoffs: -35.10 -35.10 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98416126 Reference Title: Cobalamin (vitamin B12) biosynthesis: identification and Reference Title: characterization of a Bacillus megaterium cobl operon. Reference Author: Raux E, Lanois A, Warren MJ, Rambach A, Thermes C; Reference Location: Biochem J 1998;335:159-166. Database Reference: INTERPRO; IPR002762; Database reference: PFAMB; PB040604; Database reference: PFAMB; PB040610; Database reference: PFAMB; PB041575; Comment: The function of CbiX is uncertain, however it is found Comment: in cobalamin biosynthesis operons and so may have a Comment: related function. Some CbiX proteins contain a striking Comment: histidine-rich region at their C-terminus, which suggests Comment: that it might be involved in metal chelation [1]. Number of members: 6
cellulase	PDOC00565	Glycosyl hydrolases family 5 signature	The microbial degradation of cellulose and xylans requires several types of enzymes such as endoglucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91) (exoglucanases), or xylanases (EC 3.2.1.8) [1,2]. Fungi and bacteria produces a spectrum of cellulolytic enzymes (cellulases) and xylanases which, on the basis of sequence similarities, can be classified into families. One of these families is known as the cellulase family A [3] or as the glycosyl hydrolases family 5 [4,E1]. The enzymes which are currently known to belong to this family are listed below. - Endoglucanases from various species and strains of Bacillus. - Butyrivibrio fibrisolvens endoglucanases 1 (end1) and A (celA). - Caldocellum saccharolyticum bifunctional endoglucanase/exoglucanase (celB). This protein consists of two domains; it is the C-terminal domain, which has endoglucanase activity, which belongs to this family. - Clostridium acetobutylicum endoglucanase (eglA). - Clostridium cellulolyticum endoglucanases A (celcA) and D (celcD). - Clostridium cellulovorans endoglucanase B (engB) and D (engD). - Clostridium thermocellum endoglucanases B (celB), C (celC), E (celE), G (celG) and H (celH). - Erwinia chrysanthemi endoglucanase Z (celZ). - Fibrobacter succinogenes endoglucanase 3 (cel-3). - Pseudomonas fluorescens endoglucanase C (celC).

Pfam	Prosite	Full Name	Description
			<ul style="list-style-type: none"> - <i>Pseudomonas solanacearum</i> endoglucanase (egl). - Robillarda strain Y-20 endoglucanase I. - <i>Ruminococcus albus</i> endoglucanases I (EG-I), A (celA), and B (celB). - <i>Ruminococcus flavefaciens</i> cellodextrinase A (celA). - <i>Ruminococcus flavefaciens</i> endoglucanase E (celE). - <i>Streptomyces lividans</i> endoglucanase. - <i>Thermomonospora fusca</i> endoglucanase E-5 (celE). - <i>Trichoderma reesei</i> endoglucanase II (EGLII). - <i>Xanthomonas campestris</i> endoglucanase (engxcA). <p>As well as:</p> <ul style="list-style-type: none"> - Baker's yeast glucan 1,3-beta-glucosidase I/II (EC 3.2.1.58) (EXG1). - Baker's yeast glucan 1,3-beta-glucosidase 2 (EC 3.2.1.58) (EXG2). - Baker's yeast sporulation-specific glucan 1,3-beta-glucosidase (SPR1). - <i>Caldocellum saccharolyticum</i> beta-mannanase (EC 3.2.1.78) (manA). - Yeast hypothetical protein YBR056w. - Yeast hypothetical protein YIR007w. <p>One of the conserved regions in these enzymes contains a conserved glutamic acid residue which is potentially involved [5] in the catalytic mechanism.</p> <p>We use this region as a signature pattern.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [LIV]-[LIVMFYWGA](2)-[DNEQG]-[LIVMGST]-x-N-E-[PV]-[RHDNSTLIVFY] [E is a putative active site residue] Sequences known to belong to this class detected by the pattern ALL, except for Robillarda Y-20 endoglucanase I whose sequence is known to be incorrect and yeast YBR056w. Other sequence(s) detected in SWISS-PROT 22. Expert(s) to contact by email Henrissat B. bernie@afmb.cnrs-mrs.fr</p> <p>Last update November 1997 / Pattern and text revised.</p> <p>References [1] Beguin P. Annu. Rev. Microbiol. 44:219-248(1990).</p> <p>[2] Gilkes N.R., Henrissat B., Kilburn D.G., Miller R.C. Jr., Warren R.A.J. Microbiol. Rev. 55:303-315(1991).</p> <p>[3] Henrissat B., Claeysens M., Tomme P., Lemesle L., Mornon J.-P. Gene 81:83-95(1989).</p> <p>[4] Henrissat B. Biochem. J. 280:309-316(1991).</p> <p>[5] Py B., Bortoli-German I., Haiech J., Chippaux M., Barras F. Protein Eng. 4:325-333(1991).</p> <p>[E1] http://www.expasy.ch/cgi-bin/lists?glycosid.txt</p>
CH	PDOC00019	Actinin-type actin-binding domain signatures	Alpha-actinin is a F-actin cross-linking protein which is thought to anchor actin to a variety of intracellular structures [1]. The actin-binding

DOCTOT " 08668960

Pfam	Prosite	Full Name	Description
			<p>domain of alpha-actinin seems to reside in the first 250 residues of the protein. A similar actin-binding domain has been found in the N-terminal region of many different actin-binding proteins [2,3]:</p> <ul style="list-style-type: none"> - In the beta chain of spectrin (or fodrin). - In dystrophin, the protein defective in Duchenne muscular dystrophy (DMD) and which may play a role in anchoring the cytoskeleton to the plasma membrane. - In the slime mold gelation factor (or ABP-120). - In actin-binding protein ABP-280 (or filamin), a protein that link actin filaments to membrane glycoproteins. - In fimbrin (or plastin), an actin-bundling protein. Fimbrin differs from the above proteins in that it contains two tandem copies of the actin-binding domain and that these copies are located in the C-terminal part of the protein. <p>We selected two conserved regions as signature patterns for this type of domain. The first of this region is located at the beginning of the domain, while the second one is located in the central section and has been shown to be essential for the binding of actin.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [EQ]-x(2)-[ATV]-[FY]-x(2)-W-x-N Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 25.</p> <p>Consensus pattern [LIVM]-x-[SGN]-[LIVM]-[DAGHE]-[SAG]-x-[DNEAG]-[LIVM]-x-[DEAG]-x(4)-[LIVM]-x-[LM]-[SAG]-[LIVM]-[LIVMT]-W-x-[LIVM](2) Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update November 1997 / Patterns and text revised. References [1] Schleicher M., Andre E., Harmann A., Noegel A.A. Dev. Genet. 9:521-530(1988).</p> <p>[2] Matsudaira P. Trends Biochem. Sci. 16:87-92(1991).</p> <p>[3] Dubreuil R.R. BioEssays 13:219-226(1991).</p>
chitinase_2	PDOC00839	Chitinases family 18 active site	<p>Chitinases (EC 3.2.1.14) [1] are enzymes that catalyze the hydrolysis of the beta-1,4-N-acetyl-D-glucosamine linkages in chitin polymers. From the view point of sequence similarity chitinases belong to either family 18 or 19 in the classification of glycosyl hydrolases [2,E1]. Chitinases of family 18 (also known as classes III or V) groups a variety of proteins:</p> <p>a) Chitinases from:</p>

006101" 08668960

Pfam	Prosite	Full Name	Description
			[E1] http://www.expasy.ch/cgi-bi/lists?glycosid.txt
Choline_kinase		Choline/ethanolamine kinase	<p>Accession number: PF01633</p> <p>Definition: Choline/ethanolamine kinase</p> <p>Author: Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_1165 (release 4.1)</p> <p>Gathering cutoffs: 25 25</p> <p>Trusted cutoffs: 242.90 242.90</p> <p>Noise cutoffs: -85.90 -85.90</p> <p>HMM build command line: hmmbuild -F HMM SEED</p> <p>HMM build command line: hmmcalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 98175949</p> <p>Reference Title: Expression, purification, and characterization of choline</p> <p>Reference Title: kinase, product of the CKI gene from <i>Saccharomyces cerevisiae</i>.</p> <p>Reference Author: Kim KH, Voelker DR, Flocco MT, Carman GM;</p> <p>Reference Location: J Biol Chem 1998;273:6844-6852.</p> <p>Database Reference: INTERPRO; IPR002573;</p> <p>Comment: Choline kinase catalyses the committed step in the synthesis of</p> <p>Comment: phosphatidylcholine by the CDP-choline pathway [1].</p> <p>Number of members: 22</p>
Chorion		Chorion protein	<p>Accession number: PF01723</p> <p>Definition: Chorion protein</p> <p>Author: Bashton M, Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_1914 (release 4.1)</p> <p>Gathering cutoffs: -46 -46</p> <p>Trusted cutoffs: -43.70 -43.70</p> <p>Noise cutoffs: -49.00 -49.00</p> <p>HMM build command line: hmmbuild -F HMM SEED</p> <p>HMM build command line: hmmcalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 95333194</p> <p>Reference Title: Sequence analysis of a small early chorion gene subfamily</p> <p>Reference Title: interspersed within the late gene locus in <i>Bombyx mori</i>.</p> <p>Reference Author: Kravariti L, Lecanidou R, Rodakis GC;</p> <p>Reference Location: J Mol Evol 1995;41:24-33.</p> <p>Reference Number: [2]</p> <p>Reference Medline: 86313609</p> <p>Reference Title: Evolution of the silk moth chorion gene superfamily: gene</p> <p>Reference Title: families CA and CB.</p> <p>Reference Author: Lecanidou R, Rodakis GC, Eickbush TH, Kafatos FC;</p> <p>Reference Location: Proc Natl Acad Sci U S A 1986;83:6514-6518.</p> <p>Database Reference: INTERPRO; IPR002635;</p> <p>Database reference: PFAM; PB009425;</p> <p>Comment: This family consists of the chorion superfamily proteins</p> <p>Comment: classes A, B, CA, CB and high-cysteine HCB from silk,</p> <p>Comment: gypsy and polyphemus moths.</p> <p>Comment: The chorion proteins make up the moths egg shell a complex</p> <p>Comment: extracellular structure [2].</p> <p>Number of members: 35</p>
Chorismate_mut		Chorismate mutase	<p>Accession number: PF01817</p> <p>Definition: Chorismate mutase</p> <p>Author: Bateman A</p> <p>Alignment method of seed: Manual</p>

Pfam	Prosite	Full Name	Description
			<p>Source of seed members: PSI-BLAST 1ecm</p> <p>Gathering cutoffs: 5 5</p> <p>Trusted cutoffs: 5.10 5.10</p> <p>Noise cutoffs: -19.90 -19.90</p> <p>HMM build command line: hmmbuild HMM SEED</p> <p>HMM build command line: hmmcalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 95062155</p> <p>Reference Title: The crystal structure of allosteric chorismate mutase at 2.2-A resolution.</p> <p>Reference Author: Xue Y, Lipscomb WN, Graf R, Schnappauf G, Braus G;</p> <p>Reference Location: Proc Natl Acad Sci U S A 1994;91:10814-10818.</p> <p>Reference Number: [2]</p> <p>Reference Medline: 98307941</p> <p>Reference Title: Tyrosine and tryptophan act through the same binding site</p> <p>Reference Title: at the dimer interface of yeast chorismate mutase.</p> <p>Reference Author: Schnappauf G, Krappmann S, Braus GH;</p> <p>Reference Location: J Biol Chem 1998;273:17012-17017.</p> <p>Reference Number: [3]</p> <p>Reference Medline: 98165805</p> <p>Reference Title: Chorismate mutase-prephenate dehydratase from Escherichia coli. Study of catalytic and regulatory domains using genetically engineered proteins.</p> <p>Reference Author: Zhang S, Pohnert G, Kongsaree P, Wilson DB, Clardy J,</p> <p>Reference Author: Ganem B;</p> <p>Reference Location: J Biol Chem 1998;273:6248-6253.</p> <p>Database Reference: SCOP; 1csm; fa; [SCOP-USA][CATH-PDBSUM]</p> <p>Database Reference INTERPRO; IPR002701;</p> <p>Database Reference PDB; 1ecm B; 6; 89;</p> <p>Database Reference PDB; 1ecm A; 5; 89;</p> <p>Database Reference PDB; 1csm A; 133; 162;</p> <p>Database Reference PDB; 3csm A; 133; 243;</p> <p>Database Reference PDB; 3csm B; 133; 243;</p> <p>Database Reference PDB; 4csm A; 133; 243;</p> <p>Database Reference PDB; 4csm B; 133; 243;</p> <p>Database Reference PDB; 5csm A; 133; 243;</p> <p>Database Reference PDB; 2csm A; 133; 246;</p> <p>Comment: Chorismate mutase EC:5.4.99.5 catalyses the conversion of</p> <p>Comment: chorismate to prephenate in the pathway of tyrosine and</p> <p>Comment: phenylalanine biosynthesis. This enzyme is negatively</p> <p>Comment: regulated by tyrosine, tryptophan and phenylalanine [2,3].</p> <p>Number of members: 28</p>
CN_hydrolase	PDOC00712; PDOC00943	Nitrilases / cyanide hydratase signatures; Uncharacterized protein family UPF0012 signature	<p>Nitrilases (EC 3.5.5.1) are enzymes that convert nitriles into their corresponding acids and ammonia. They are widespread in microbes as well as in plants where they convert indole-3-acetonitrile to the hormone indole-3-acetic acid. A conserved cysteine has been shown [1,2] to be essential for enzyme activity; it seems to be involved in a nucleophilic attack on the nitrile carbon atom.</p> <p>Cyanide hydratase (EC 4.2.1.66) converts HCN to formamide. In phytopathogenic fungi, it is used to avoid the toxic effect of cyanide released by wounded plants [3]. The sequence of cyanide hydrolase is evolutionary related to that</p>

09689980 1013000

'C': conserved cysteine involved in a disulfide bond.
'*': position of the patterns.

Pfam	Prosite	Full Name	Description
			<p>Apart from regions around some of the histidine heme ligands, there are a few conserved regions in the sequence of b/b6. The best conserved of these regions includes an invariant P-E-W triplet which lies in the loop that separates the fifth and sixth transmembrane segments. It seems to be important for electron transfer at the ubiquinone redox site - called Qz or Qo (where o stands for outside) - located on the outer side of the membrane.</p> <p>A schematic representation of the structure of cytochrome b/b6 is shown below.</p> <pre> +---Fe-b562---+ +---Fe-b566--- + </pre> <p>xxxxxxxxxxxxHxHxxxxxxxxxxxxHxHxxxxxxxxxxxxPEWxxxxxxxxxxxx xxxxx</p> <p><-----Cytochrome-b-----> <----Cytochrome-b6-petB-----><--Cytochrome-b6-petD-----></p> <p>We developed two signature patterns for cytochrome b/b6. The first includes the first conserved histidine of b/b6, which is a heme b562 ligand; the second includes the conserved PEW triplet.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [DENQ]-x(3)-G-[FYWMQ]-x-[LIVMF]-R-x(2)-H [H is a heme b562 ligand] Sequences known to belong to this class detected by the pattern ALL, except for 5 sequences. Other sequence(s) detected in SWISS-PROT 15.</p> <p>Consensus pattern P-[DE]-W-[FY]-[LFY](2) Sequences known to belong to this class detected by the pattern ALL, except for <i>Odocoileus hemionus</i> (mule deer) and <i>Paramecium tetraurelia</i> cytochrome b. Other sequence(s) detected in SWISS-PROT 1. Last update November 1995 / Patterns and text revised. References [1] Howell N. J. Mol. Evol. 29:157-169(1989). [2] Esposti M.D., de Vries S., Crimi M., Ghelli A., Patarnello T., Meyer A. Biochim. Biophys. Acta 1143:243-271(1993).</p>
cytochrome_b_N	PDOC00171	Cytochrome b/b6 signatures	<p>In the mitochondrion of eukaryotes and in aerobic prokaryotes, cytochrome b is a component of respiratory chain complex III (EC 1.10.2.2) - also known as the bc1 complex or ubiquinol-cytochrome c reductase. In plant chloroplasts and cyanobacteria, there is a analogous protein, cytochrome b6, a component of the plastoquinone-plastocyanin reductase (EC 1.10.99.1), also known as the b6f complex.</p> <p>Cytochrome b/b6 [1,2] is an integral membrane protein of approximately 400 amino acid residues that probably has 8 transmembrane segments. In plants and</p>

09689980 "101300

Pfam	Prosite	Full Name	Description
			<p>into the endoplasmic reticulum and the nucleus.</p> <ul style="list-style-type: none"> - Yeast protein SIS1, required for nuclear migration during mitosis. - Yeast protein CAJ1. - Yeast hypothetical protein YFR041c. - Yeast hypothetical protein YIR004w. - Yeast hypothetical protein YJL162c. - Plasmodium falciparum ring-infected erythrocyte surface antigen (RESA). <p>RESA, whose function is not known, is associated with the membrane skeleton of newly invaded erythrocytes.</p> <ul style="list-style-type: none"> - Human HDJ1. - Human HSJ1, a neuronal protein. - Drosophila cysteine-string protein (csp). <p>We developed a signature pattern for the 'J' domain, based on conserved positions in the C-terminal half of this domain. We also developed a pattern for the 'CRR' domain, based on the first two copies of that motif. We also developed a profile for the 'J' domain.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [FY]-x(2)-[LIVMA]-x(3)-[FYWHNT]-[DENQSA]-x-L-x-[DN]-x(3)-[KR]-x(2)-[FYI] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 5.</p> <p>Consensus pattern C-[DEGSTHKR]-x-C-x-G-x-[GK]-[AGSDM]-x(2)-[GSNKR]-x(4,6)-C-x(2,3)-C-x-G-x-G Sequences known to belong to this class detected by the pattern ALL, except for yeast XDJ1. Other sequence(s) detected in SWISS-PROT 8.</p> <p>Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so. Expert(s) to contact by email Kelley W. kelley@medecine.unige.ch</p> <p>Last update July 1998 / Patterns and text revised. References [1] Cyr D.M., Langer T., Douglas M.G. Trends Biochem. Sci. 19:176-181(1994).</p> <p>[2] Bork P., Sander C., Valencia A., Bukau B. Trends Biochem. Sci. 17:129-129(1992).</p> <p>[3] Ueguchi C., Kaneda M., Yamada H., Mizuno T. Proc. Natl. Acad. Sci. U.S.A. 91:1054-1058(1994).</p>
dNK		Deoxynucleoside kinase	<p>Accession number: PF01712 Definition: Deoxynucleoside kinase Author: Bashon M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1744 (release 4.1) Gathering cutoffs: 25 25 Trusted cutoffs: 47.50 47.50</p>

09669980-101300

Pfam	Prosite	Full Name	Description
			<p>Noise cutoffs: -5.40 -5.40 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97236800 Reference Title: Cloning of the cDNA and chromosome localization of the gene Reference Title: for human thymidine kinase 2. Reference Author: Johansson M, Karlsson A; Reference Location: J Biol Chem 1997;272:8454-8458. Reference Number: [2] Reference Medline: 96293511 Reference Title: Cloning and expression of human deoxyguanosine kinase cDNA. Reference Author: Johansson M, Karlsson A; Reference Location: Proc Natl Acad Sci U S A 1996;93:7258-7262. Database Reference INTERPRO; IPR002624; Comment: This family consists of various deoxynucleoside kinases Comment: cytidine EC:2.7.1.74, guanosine EC:2.7.1.113, adenosine EC:2.7.1.76 Comment: and thymidine kinase EC:2.7.1.21 (which also phosphorylates deoxyuridine Comment: and deoxycytosine.) These enzymes catalyse the production of Comment: deoxynucleotide 5'-monophosphate from a deoxynucleoside. Comment: Using ATP and yielding ADP in the process. Number of members: 20</p>
DSL		Delta serrate ligand	<p>Accession number: PF01414 Definition: Delta serrate ligand Author: Ponting CP, Schultz J, Bork P Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 25 25 Trusted cutoffs: 43.00 43.00 Noise cutoffs: 3.40 3.40 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96125168 Reference Title: Interchangeability of Caenorhabditis elegans DSL proteins Reference Title: and intrinsic signalling activity of their extracellular Reference Title: domains in vivo. Reference Author: Fitzgerald K, Greenwald I; Reference Location: Development 1995;121:4275-4282. Reference Number: [2] Reference Medline: 92034990 Reference Title: Specific EGF repeats of Notch mediate interactions with Reference Title: Delta and Serrate: implications for Notch as a Reference Title: multifunctional receptor. Reference Author: Rebay I, Fleming RJ, Fehon RG, Cherbas L, Cherbas P, Reference Author: Artavanis-Tsakonas S; Reference Location: Cell 1991;67:687-699. Reference Number: [3] Reference Medline: 95232495 Reference Title: Notch signaling. Reference Author: Artavanis-Tsakonas S, Matsuno K, Fortini ME; Reference Location: Science 1995;268:225-232. Database reference: SMART; DSL; Database Reference INTERPRO; IPR001774; Number of members: 30</p>
DUF125		Integral membrane protein DUF125	<p>Accession number: PF01988 Definition: Integral membrane protein DUF125</p>

005707 " 08668960

Pfam	Prosite	Full Name	Description
			<p>Consensus pattern H-x-l-x-G-[KR]-x-F-[GA]-S-x-V-[ST]-[HY]-E</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update November 1997 / First entry.</p> <p>References</p> <p>[1] King S.M., Patel-King R.S. J. Biol. Chem. 270:11445-11452(1995).</p> <p>[2] Dick T., Ray K., Salz H.K., Chia W. Mol. Cell. Biol. 16:1966-1977(1996).</p>
elF5_elF2B		Domain found in IF2B/IF5	<p>Accession number: PF01873</p> <p>Definition: Domain found in IF2B/IF5</p> <p>Author: Enright A, Ouzounis C, Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Enright A</p> <p>Gathering cutoffs: 25 25</p> <p>Trusted cutoffs: 233.00 233.00</p> <p>Noise cutoffs: -56.10 -56.10</p> <p>HMM build command line: hmmbuild -F HMM SEED</p> <p>HMM build command line: hmmbuild --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 96060092</p> <p>Reference Title: Multidomain organization of eukaryotic guanine nucleotide exchange translation initiation factor eIF-2B subunits</p> <p>Reference Title: revealed by analysis of conserved sequence motifs.</p> <p>Reference Author: Koonin EV;</p> <p>Reference Location: Protein Sci 1995;4:1608-1617.</p> <p>Database Reference: INTERPRO; IPR002735;</p> <p>Comment: This family includes the N terminus of eIF-5 Swiss:P55010, and</p> <p>Comment: the C terminus of eIF-2 beta Swiss:P20042.</p> <p>This region</p> <p>Comment: corresponds to the whole of the archaeobacterial eIF-2 beta</p> <p>Comment: homolog. The region contains a putative zinc binding C4 finger.</p> <p>Number of members: 20</p>
elF6		eIF-6 family	<p>This family comprises members exhibiting sequence identity to the eukaryotic translation initiation factor 6. Some members of this family are implicated in protein biosynthesis as a translation initiation factor by binding to the 60s ribosomal subunit and preventing its association with the 40s ribosomal subunit to form the 80s initiation complex. Such activity can play a role in maximal polysome formation and plays an important role in determining free 60s ribosomal subunit content. Polypeptides in this family can optimize amino acid and nitrogen content in a desired cell or organism. References describing eif6 family members and their biological activities include, for example, the following: Adams et al., Science 87:2185-2195(2000); Wood et al., J. Biol. Chem. 274:11653-11659(1999); and Si et al., Mol. Cell. Biol. 19:1416-1426(1999).</p>
ER	PDOC00992	Enhancer of rudimentary signature	<p>The Drosophila protein 'enhancer of rudimentary' (gene (e(r)) is a small protein of 104 residues whose function is not yet clear. From an evolutionary point of view, it is highly conserved [1] and has been found to exist in probably all multicellular eukaryotic organisms. It has been proposed that this protein plays a role in the cell cycle.</p> <p>As as signaure pattern, we selected a conserved region in the central part of</p>

09669980 101300

Pfam	Prosite	Full Name	Description
			<p>Sequences known to belong to this class detected by the pattern ALL, except for Mycoplasma pneumoniae aldolase. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Consensus pattern [LIVM]-E-x-E-[LIVM]-G-x(2)-[GM]-[GSTA]-x-E Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update December 1999 / Pattern and text revised. References [1] Perham R.N. Biochem. Soc. Trans. 18:185-187(1990).</p> <p>[2] Marsh J.J., Lebherz H.G. Trends Biochem. Sci. 17:110-113(1992).</p> <p>[3] von der Osten C.H., Barbas C.F. III, Wong C.-H., Sinskey A.J. Mol. Microbiol. 3:1625-1637(1989).</p> <p>[4] Berry A., Marshall K.E. FEBS Lett. 318:11-16(1993).</p>
FAA_hydrolase		Fumarylacetoacetate (FAA) hydrolase family	<p>Accession number: PF01557 Definition: Fumarylacetoacetate (FAA) hydrolase family Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_641 (release 4.0) Gathering cutoffs: 25 25 Trusted cutoffs: 42.10 42.10 Noise cutoffs: -93.10 -93.10 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97255958 Reference Title: Mutations in the fumarylacetoacetate hydrolase gene causing hereditary tyrosinemia type I: overview. Reference Author: St-Louis M, Tanguay RM; Reference Location: Hum Mutat 1997;9:291-299. Reference Number: [2] Reference Medline: 96125235 Reference Title: Molecular characterization of the 4-hydroxyphenylacetate catabolic pathway of Escherichia coli W: engineering a mobile aromatic degradative cluster. Reference Author: Prieto MA, Diaz E, Garcia JL; Reference Location: J Bacteriol 1996;178:111-120. Reference Number: [3] Reference Medline: 96016123 Reference Title: Fungal metabolic model for human type I hereditary tyrosinaemia. Reference Author: Fernandez-Canon JM, Penalva MA; Reference Location: Proc Natl Acad Sci U S A 1995;92:9132-9136. Reference Number: [4] Reference Medline: 94039092 Reference Title: Purification, nucleotide sequence and some properties of a bifunctional isomerase/decarboxylase from the homoprotocatechuate degradative pathway of Escherichia coli C. Reference Author: Roper DI, Cooper RA; Reference Location: Eur J Biochem 1993;217:575-580. Database reference: MIM; 276700; Database Reference: INTERPRO; IPR002529;</p>

Pfam	Prosite	Full Name	Description
			<p>This family consists of fumarylacetoacetate (FAA) hydrolase, Comment: or fumarylacetoacetate hydrolase (FAH) and it also includes Comment: HHDD isomerase/OPET decarboxylase from E. coli strain W. Comment: FAA is the last enzyme in the tyrosine catabolic pathway, it hydrolyses Comment: fumarylacetoacetate into fumarate and acetoacetate which then join the Comment: citric acid cycle [1]. Mutations in FAA cause type I tyrosinemia in humans Comment: this is an inherited disorder mainly affecting the liver leading to Comment: liver cirrhosis, hepatocellular carcinoma, renal tubular damages and Comment: neurologic crises amongst other symptoms [1]. The enzymatic defect causes Comment: the toxic accumulation of phenylalanine/tyrosine catabolites [3]. Comment: The E. coli W enzyme HHDD isomerase/OPET decarboxylase contains two Comment: copies of this domain and functions in fourth and fifth steps of the Comment: homoprotocatechuate pathway; Comment: here it decarboxylates OPET to HHDD and isomerizes this to OHED. Comment: The final products of this pathway are pyruvic acid and succinic Comment: semialdehyde. Number of members: 33</p>
FAD_binding		FAD binding domain	<p>Accession number: PF00667 Definition: FAD binding domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_180 (release 2.1) Gathering cutoffs: 16.8 16.8 Trusted cutoffs: 24.60 16.80 Noise cutoffs: 13.50 15.90 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 95386502 Reference Title: The flavin reductase activity of the flavoprotein component Reference Title: of sulfite reductase from Escherichia coli. A new model for Reference Title: the protein structure. Reference Author: Eschenbrenner M, Coves J, Fontecave M; Reference Location: J Biol Chem 1995;270:20550-20555. Reference Number: [2] Reference Medline: 96049560 Reference Title: NADPH-sulfite reductase flavoprotein from Escherichia coli: Reference Title: contribution to the flavin content and subunit interaction. Reference Author: Eschenbrenner M, Coves J, Fontecave M; Reference Location: FEBS Lett 1995;374:82-84. Reference Number: [3] Reference Medline: 94360001 Reference Title: Dissection of NADPH-cytochrome P450 oxidoreductase into Reference Title: distinct functional domains. Reference Author: Smith GC, Tew DG, Wolf CR; Reference Location: Proc Natl Acad Sci U S A 1994;91:8710- 8714. Reference Number: [4] Reference Medline: 97385116 Reference Title: Three-dimensional structure of NADPH- cytochrome P450 Reference Title: reductase: prototype for FMN- and FAD- containing enzymes. Reference Author: Wang M, Roberts DL, Paschke R, Shea</p>

09689980-101300

Pfam	Prosite	Full Name	Description
			<p>TM, Masters BS, Kim JJ; Reference Location: Proc Natl Acad Sci U S A 1997;94:8411-8416. Database Reference: SCOP; 1amo; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR001709; Database Reference PDB; 1amo A; 274; 493; Database Reference PDB; 1amo B; 274; 493; Database Reference PDB; 1quf ; 77; 120; Database reference: PFAMB; PB001390; Comment: This domain is found in sulfite reductase, NADPH cytochrome P450 Comment: reductase and Nitric oxide synthase. Number of members: 87</p>
FAD_binding_3		FAD binding domain	<p>Accession number: PF01494 Definition: FAD binding domain Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_549 (release 4.0) Gathering cutoffs: -7 -7 Trusted cutoffs: -6.20 -6.20 Noise cutoffs: -7.90 -7.90 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 93028353 Reference Title: Crystal structure of the reduced form of p-hydroxybenzoate Reference Title: hydroxylase refined at 2.3A resolution. Reference Author: Schreuder HA, van der Laan JM, Swarte MB, Kalk KH, Hol WG, Reference Author: Drenth J; Reference Location: Proteins 1992;14:178-190. Database Reference: SCOP; 2phh; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR002938; Database Reference PDB; 1pxa ; 5; 35; Database Reference PDB; 1bf3 ; 5; 139; Database Reference PDB; 1bgj ; 5; 139; Database Reference PDB; 1bgn ; 5; 139; Database Reference PDB; 1bkw ; 5; 139; Database Reference PDB; 1cc4 A; 5; 139; Database Reference PDB; 1cc6 A; 5; 139; Database Reference PDB; 1cj2 A; 5; 139; Database Reference PDB; 1pbb ; 5; 139; Database Reference PDB; 1pbc ; 5; 139; Database Reference PDB; 1pbd ; 5; 139; Database Reference PDB; 1pbe ; 5; 139; Database Reference PDB; 1pbf ; 5; 139; Database Reference PDB; 1pdh ; 5; 139; Database Reference PDB; 2phh ; 5; 139; Database Reference PDB; 1cj3 A; 5; 139; Database Reference PDB; 1cj4 A; 5; 139; Database Reference PDB; 1phh ; 5; 139; Database Reference PDB; 1d7l A; 5; 139; Database Reference PDB; 1dob ; 5; 139; Database Reference PDB; 1doc ; 5; 139; Database Reference PDB; 1dod ; 5; 139; Database Reference PDB; 1doe ; 5; 139; Database Reference PDB; 1ius ; 5; 139; Database Reference PDB; 1iut ; 5; 139; Database Reference PDB; 1iuu ; 5; 139; Database Reference PDB; 1iuv ; 5; 139; Database Reference PDB; 1iuw ; 5; 139; Database Reference PDB; 1iux ; 5; 139; Database Reference PDB; 1foh A; 10; 151; Database Reference PDB; 1foh D; 10; 151; Database Reference PDB; 1foh B; 10; 151; Database Reference PDB; 1foh C; 10; 151; Database reference: PFAMB; PB040546; Comment: This domain is involved in FAD binding in a number of enzymes. Number of members: 52</p>

Pfam	Prosite	Full Name	Description
		region signature	<p>into several subgroups depending upon the physiological nature of the iron sulfur cluster(s) and according to sequence similarities. One of these subgroups are the 2Fe-2S ferredoxins, which are proteins or domains of around one hundred amino acid residues that bind a single 2Fe-2S iron-sulfur cluster.</p> <p>The proteins that are known [2] to belong to this family are listed below.</p> <ul style="list-style-type: none"> - Ferredoxin from photosynthetic organisms; namely plants and algae where it is located in the chloroplast or cyanelle; and cyanobacteria. - Ferredoxin from archaeobacteria of the Halobacterium genus. - Ferredoxin IV (gene pftA) and V (gene fdxD) from Rhodobacter capsulatus. - Ferredoxin in the toluene degradation operon (gene xylT) and naphthalene degradation operon (gene nahT) of Pseudomonas putida. - Hypothetical Escherichia coli protein yfaE. - The N-terminal domain of the bifunctional ferredoxin/ferredoxin reductase electron transfer component of the benzoate 1,2-dioxygenase complex (gene benC) from Acinetobacter calcoaceticus, the toluene 4-monooxygenase complex (gene tmoF), the toluate 1,2-dioxygenase system (gene xylZ), and the xylene monooxygenase system (gene xylA) from Pseudomonas. - The N-terminal domain of phenol hydroxylase protein p5 (gene dmpP) from Pseudomonas Putida. - The N-terminal domain of methane monooxygenase component C (gene mmoC) from Methylococcus capsulatus . - The C-terminal domain of the vanillate degradation pathway protein vanB in a Pseudomonas species. - The N-terminal domain of bacterial fumarate reductase iron-sulfur protein (gene frdB). - The N-terminal domain of CDP-6-deoxy-3,4-glucoseen reductase (gene ascD) from Yersinia pseudotuberculosis. - The central domain of eukaryotic succinate dehydrogenase (ubiquinone) iron-sulfur protein. - The N-terminal domain of eukaryotic xanthine dehydrogenase. - The N-terminal domain of eukaryotic aldehyde oxidase. <p>In the 2Fe-2S ferredoxins, four cysteine residues bind the iron-sulfur cluster. Three of these cysteines are clustered together in the same region of the protein. Our signature pattern spans that iron-sulfur binding region.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern C-[C]-[C]-[GA]-[C]-C-[GAST]-{CPDEKRHFYW}-C [The three C's are 2Fe-2S ligands] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 15.</p> <p>Note in addition to the proteins listed above there are a number of other ferredoxin-like proteins that bind a 2Fe-2S cluster but which do not seem to be evolutionary related to this family. Among them are the ferredoxins from the adrenodoxin family (see <PDOC00642>) as well as the bacterial aromatic dioxygenase</p>

0966980 "101300

Pfam

Prosite	F

Description
systems ferredoxin-like proteins such as bnzC, ndoA, and todB.
Last update
November 1997 / Text revised.
References
[1]
Meyer J.
Trends Ecol. Evol. 3:222-226(1988).
[2]
Harayama S., Polissi A., Reikik M.
FEBS Lett. 285:85-88(1991).
Ferredoxins [1] are a group of iron-sulfur proteins which mediate electron transfer in a wide variety of metabolic reactions. Ferredoxins can be divided into several subgroups depending upon the physiological nature of the iron sulfur cluster(s) and according to sequence similarities. One family of ferredoxins groups together the following proteins that all bind a single 2Fe-2S iron-sulfur cluster:
- Adrenodoxin (ADX) (adrenal ferredoxin), a vertebrate mitochondrial protein which transfers electrons from adrenodoxin reductase to cytochrome P450_{scc}, which is involved in cholesterol side chain cleavage.
- Putidaredoxin (PTX), a Pseudomonas putida protein which transfers electrons from putidaredoxin reductase to cytochrome P450_{cam}, which is involved in the oxidation of camphor.
- Terpredoxin [2], a Pseudomonas protein which transfers electrons from terpredoxin reductase to cytochrome P450_{terp}, which is involved in the oxidation of alpha-terpineol.
- Rhodocoxin [3], a Rhodococcus protein which transfers electrons from rhodocoxin reductase to cytochrome CYP116 (thcB), which is involved in the degradation of thiocarbamate herbicides.
- Escherichia coli ferredoxin (gene fdx) [4] whose exact function is not yet known.
- Rhodobacter capsulatus ferredoxin VI [5], which may transfer electrons to a yet uncharacterized oxygenase.
- Caulobacter crescentus ferredoxin (gene fdxB) [6].
In these proteins, four cysteine residues bind the iron-sulfur cluster. Three of these cysteines are clustered together in the same region of the protein.
Our signature pattern spans that iron-sulfur binding region.
Description of pattern(s) and/or profile(s)
Consensus pattern C-x(2)-[STAQ]-x-[STAMV]-C-[STA]-T-C-[HR]
[The three C's are 2Fe-2S ligands]
Sequences known to belong to this class detected by the pattern ALL.
Other sequence(s) detected in SWISS-PROT 1.
Last update
November 1995 / Pattern and text revised.
EMBL/Genbank: X51607. **References**
[1]
Meyer J.
Trends Ecol. Evol. 3:222-226(1988).

Pfam	Prosite	Full Name	Description
			<p>(gp91-phox), ferric reductase</p> <p>Comment: transmembrane component in yeast and respiratory burst oxidase from</p> <p>Comment: mouse-ear cress.</p> <p>Comment: This may be a family of flavocytochromes capable of moving electrons</p> <p>Comment: across the plasma membrane [1].</p> <p>Comment: The Frp1 protein Swiss:Q04800 from S. pombe is a ferric reductase</p> <p>Comment: component and is required for cell surface ferric reductase activity,</p> <p>Comment: mutants in frp1 are deficient in ferric iron uptake [1].</p> <p>Comment: Cytochrome B-245 heavy chain</p> <p>Swiss:P04839 is a FAD-dependent</p> <p>Comment: dehydrogenase it is also has electron transferase activity which reduces</p> <p>Comment: molecular oxygen to superoxide anion, a precursor in the production of</p> <p>Comment: microbicidal oxidants [2].</p> <p>Comment: Mutations in the sequence of cytochrome B-245 heavy chain (gp91-phox)</p> <p>Comment: lead to the X-linked chronic granulomatous disease. The bacteriocidal</p> <p>Comment: ability of phagocytic cells is reduced and is characterised by the</p> <p>Comment: absence of a functional plasma membrane associated NADPH oxidase [3].</p> <p>Comment: The chronic granulomatous disease gene codes for the beta chain of</p> <p>Comment: cytochrome B-245 and cytochrome B-245 is missing from patients with</p> <p>Comment: the disease [4].</p> <p>Comment: The aligned region includes a potential FAD binding domain.</p> <p>Number of members: 34</p>
Flavi_NS5		Flavivirus RNA-directed RNA polymerase	<p>Accession number: PF00972</p> <p>Definition: Flavivirus RNA-directed RNA polymerase</p> <p>Author: Finn RD, Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_200 (release 3.0)</p> <p>Gathering cutoffs: 12 12</p> <p>Trusted cutoffs: 16.00 16.00</p> <p>Noise cutoffs: 8.50 8.50</p> <p>HMM build command line: hmmbuild -f HMM SEED</p> <p>HMM build command line: hmmscalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 95159427</p> <p>Reference Title: Phylogeny of TYU, SRE, and CFA virus: different</p> <p>Reference Title: evolutionary rates in the genus Flavivirus.</p> <p>Reference Author: Marin MS, Zanotto PM, Gritsun TS, Gould EA;</p> <p>Reference Location: Virology 1995;206:1133-1139.</p> <p>Reference Number: [2]</p> <p>Reference Medline: 96182933</p> <p>Reference Title: Recombinant dengue type 1 virus NS5 protein expressed in</p> <p>Reference Title: Escherichia coli exhibits RNA-dependent</p> <p>Reference Title: activity.</p> <p>Reference Author: Tan BH, Fu J, Sugrue RJ, Yap EH, Chan YC, Tan YH;</p> <p>Reference Location: Virology 1996;216:317-325.</p> <p>Reference Number: [3]</p> <p>Reference Medline: 93224895</p> <p>Reference Title: Computer-assisted identification of a putative</p> <p>Reference Title: methyltransferase domain in NS5 protein of flaviviruses and</p> <p>Reference Title: lambda 2 protein of reovirus.</p> <p>Reference Author: Koonin EV;</p> <p>Reference Location: J Gen Virol 1993;74:733-740.</p>

Pfam	Prosite	Full Name	Description
			<p>Reference Number: [4] Reference Medline: 94094568 Reference Title: Evolution and taxonomy of positive-strand RNA viruses: Reference Title: implications of comparative analysis of amino acid sequences. Reference Author: Koonin EV, Dolja VV; Reference Location: Crit Rev Biochem Mol Biol 1993;28:375-430. Database Reference INTERPRO; IPR000208; Comment: Flaviviruses produce a polypeptide from the ssRNA genome. Comment: This protein is also known as NS5. Comment: This RNA-directed RNA polymerase possesses a number of short Comment: regions and motifs homologous to other RNA-directed RNA Comment: polymerases [2]. Number of members: 159</p>
Fork_head	PDOC00564	Fork head domain signatures and profile	<p>It has been shown [1] that some eukaryotic transcription factors contain a conserved domain of about 100 amino-acid residues, called the fork head domain (but also known as a "winged helix"), which is involved in DNA-binding [2]. Proteins known to contain this domain are listed below.</p> <ul style="list-style-type: none"> - Drosophila fork head protein (fkh). Fkh is probably a transcription factor that regulates the expression of genes involved in terminal development. - Drosophila protein crocodile (gene croc) [3], which is required for the establishment of head structures. - Drosophila proteins FD2, FD3, FD4, and FD5. - Drosophila proteins sloppy paired 1 and 2 (slp1 and slp2) involved in segmentation. - Bombyx mori silk gland factor-1 (SGF-1) which regulates transcription of the sericin-1 gene. - Mammalian transcriptional activators HNF-3-alpha, -beta, and -gamma. The HNF-3 proteins interact with the cis-acting regulatory regions of a number of liver genes. - Mammalian interleukin-enhancer binding factor (ILF). ILF binds to the purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter. ILF may be involved in both positive and negative regulation of important viral and cellular promoter elements. - Mammalian transcription factor BF-1 which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon. - Human HTLF, a protein that binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR). - Mammalian transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4. - Human AFX1 which is involved in a chromosomal translocation that causes acute leukemia. - Human FKHR which is involved in a chromosomal

05689980 101300

Pfam	Prosite	Full Name	Description
			<p>translocation that causes rhabdomyosarcoma.</p> <ul style="list-style-type: none"> - Xenopus XFKH1, a protein essential for normal axis formation. - <i>Caenorhabditis elegans</i> lin-31; involved in the regulation of vulval cell fates. - Yeast HCM1, a protein of unknown function. - Yeast FKH1. - Yeast FKH2. <p>The fork domain is highly conserved. We have developed two patterns for its detection. The first corresponds to the N-terminal section of the domain; the second is a heptapeptide located in the central section of the domain.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [KR]-P-[PTQ]-[FYLVQH]-S-[FY]-x(2)-[LIVM]-x(3,4)-[AC]-[LIM] Sequences known to belong to this class detected by the pattern ALL, except for AFX1 and FKHR. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Consensus pattern W-[QKR]-[NS]-S-[LIV]-R-H Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update November 1997 / Patterns and text revised. References [1] Weigel D., Jaechle H. Cell 63:455-456(1990).</p> <p>[2] Clark K.L., Halay E.D., Lai E., Burley S.K. Nature 364:412-420(1993).</p> <p>[3] Haecker U., Kaufmann E., Hartmann C., Juergens G., Knoechel W., Jaechle H. EMBO J. 14:5306-5317(1995).</p>
FtsJ		FtsJ cell division protein	<p>Accession number: PF01728 Definition: FtsJ cell division protein Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1791 (release 4.1) Gathering cutoffs: -38 -38 Trusted cutoffs: -20.90 -20.90 Noise cutoffs: -56.70 -56.70 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 93186701 Reference Title: The Escherichia coli FtsH protein is a prokaryotic member Reference Title: of a protein family of putative ATPases involved in Reference Title: membrane functions, cell cycle control, and gene Reference Title: expression. Reference Author: Tomoyasu T, Yuki T, Morimura S, Mori H, Yamanaka K, Niki H, Reference Author: Hiraga S, Ogura T; Reference Location: J Bacteriol 1993;175:1344-1351. Database Reference: INTERPRO; IPR002877; Database reference: PFAM; PB030182; Comment: This family consists of FtsJ from various bacterial and archaeal sources</p>

Pfam	Prosite	Full Name	Description
			<p>Comment: In <i>E. coli</i> FtsJ is not essential for growth but affects cell division [1].</p> <p>Number of members: 25</p>
FTSW_RODA_SPOVE	PDOC00352	Cell cycle proteins ftsW / rodA / spoVE signature	<p>A number of prokaryotic proteins involved in cell cycle processes have been found [1,2] to be structurally related, these proteins are:</p> <ul style="list-style-type: none"> - <i>Escherichia coli</i> and related bacteria cell division protein ftsW. This protein plays a role in the stabilization of the ftsZ ring during cell division. - <i>Escherichia coli</i> and related bacteria rod shape-determining protein rodA (or mrdB). It is required for the expression of the enzymatic activity of PBP2, which is thought to participate in the synthesis of peptidoglycan during the initiation of cell elongation. - <i>Bacillus subtilis</i> stage V sporulation protein E (spoVE). The exact function of spoVE in endospore formation is not known. - <i>Bacillus subtilis</i> hypothetical protein ylaO. - <i>Bacillus subtilis</i> hypothetical protein ywcF (ipa-42D). - <i>Cyanophora paradoxa</i> cyanelle ftsW homolog. This protein may be involved in the organelle division process. <p>All these proteins are hydrophobic integral membrane protein and contain about 400 residues. We have selected the best conserved region, which is located in the C-terminal section, as a signature pattern for these proteins.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [NV]-x(5)-[GTR]-[LIVMA]-x-P-[PTLIVM]-x-G-[LIVM]-x(3)-[LIVMFVW](2)-S-[YSA]-G-G-[STN]-[SA]</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update November 1997 / Pattern and text revised.</p> <p>References</p> <p>[1] Ikeda M., Sato T., Wachi M., Jung H.K., Ishino F., Kobayashi Y., Matsuhashi M. J. Bacteriol. 171:6375-6378(1989).</p> <p>[2] Joris B., Dive G., Henriques A., Piggot P.J., Ghuysen J.-M. Mol. Microbiol. 4:513-517(1990).</p>
Furin-like		Furin-like cysteine rich region	<p>Members of this family include receptors that mediate transmembrane signalling. These receptors can bind to a number of factors including: amphiregulin, epidermal growth factor, gp30, heparin-binding egf, insulin, insulin-like growth factor I and II, neuregulins, transforming growth factor-alpha and, and vaccinia virus growth</p> <p>Signal transduction is mediated by catalytic activity of tyrosine kinase, such as ATP + A protein tyrosine = ADP + protein tyrosine phosphate. Typically, such signal transduction have been implicated in metabolic and developmental changes, including cell fate and differentiation. Examples include instruction of follicle cells to follow a dorsal pathway of development rather than the default ventral pathway. may also bind the spitz protein.</p> <p>References describing these family members and their biological activities:</p>

Pfam	Prosite	Full Name	Description
			<p>Abbot et al., J. Biol. Chem. 267:10759-10763(1992); Araki et al., J. Biol. Chem. 262:16186-16191(1987); Aroian et al., EMBO J. 13:360-366(1994); Aroian et al., Nature 348:693-699(1990); Barbetti et al., Diabetes 41:408-415(1992); Bargmann et al., Nature 319:226-230(1986); Cama et al., J. Biol. Chem. 268:8060-8069(1993); Cama et al., J. Clin. Endocrinol. Metab. 73:894-901(1991); Carrera et al., Hum. Mol. Genet. 2:1437-1441(1993); Clifford et al., Genetics 137:531-550(1994); Coccozza et al., Diabetes 41:521-526(1992); Cooke et al., Biochem. Biophys. Res. Commun. 177:1113-1120(1991); Coussens et al., Science 230:1132-1139(1985); Dickens et al., Biochem. Biophys. Res. Commun. 186:244-250(1992); Ebina et al., Cell 40:747-758(1985); Ebina et al., Proc. Natl. Acad. Sci. U.S.A. 84:704-708(1987); Ehsani et al., Genomics 15:426-429(1993); Elbein et al., Diabetes 42:429-434(1993); Elbein, Diabetes 38:737-743(1989); Fujita-Yamaguchi et al., Protein Seq. Data Anal. 1:3-6(1987); Gullick et al., EMBO J. 11:43-48(1992); Haruta et al., Diabetes 42:1837-1844(1993); Hubbard et al., EMBO J. 16:5572-5581(1997).</p> <p>Hubbard et al., Nature 372:746-754(1994); Iwanishi et al., Diabetologia 36:414-422(1993); Kadowaki et al., J. Clin. Invest. 86:254-264(1990); Kadowaki et al., Science 240:787-790(1988); Kim et al., Diabetologia 35:261-266(1992); Klinkhamer et al., EMBO J. 8:2503-2507(1989); Kusari et al., J. Biol. Chem. 266:5260-5267(1991); Lai et al., Neuron 6:691-704(1991); Lax et al., Mol. Cell. Biol. 8:1970-1978(1988); Lebrun et al., J. Biol. Chem. 268:11272-11277(1993); Lee et al., Oncogene 8:3403-3410(1993); Lesokhin et al., Dev. Biol. 205:129-144(1999); Livneh et al., Cell 40:599-607(1985).</p> <p>Longo et al., Proc. Natl. Acad. Sci. U.S.A. 90:60-64(1993); McKeon et al., Mol. Endocrinol. 4:647-656(1990); Moller et al., J. Biol. Chem. 265:14979-14985(1990); Moller et al., Mol. Endocrinol. 4:1183-1191(1990); Odawara et al., Science 245:66-68(1989); Raz et al., Genetics 129:191-201(1991).</p> <p>Sakai et al., J. Mol. Biol. 256:548-555(1996); Schaeffer et al., Biochem. Biophys. Res. Commun. 189:650-653(1992); Schejter et al., Cell 46:1091-1101(1986); Seino et al., Biochem. Biophys. Res. Commun. 159:312-316(1989); Seino et al., Diabetes 39:123-128(1990); Semba et al., Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985); Shier et al., J. Biol. Chem. 264:14605-14608(1989); Taira et al., Science 245:63-66(1989); Tewari et al., J. Biol. Chem. 264:16238-16245(1989); Ullrich et al., Nature 313:756-761(1985).</p> <p>Ullrich et al., EMBO J. 5:2503-2512(1986); van der Vorm et al., Diabetologia 36:172-174(1993); van der Vorm et al., J. Biol. Chem. 267:66-71(1992); Wadsworth et al., Nature 314:178-180(1985); White et al., Cell 54:641-649(1988); Xu et al., J. Biol. Chem. 265:18673-18681(1990); Yamamoto et al., Nature 319:230-234(1986); and Yoshimasa et al., Science 240:784-787(1988).</p>
Galactosyl_T		Galactosyltransferase	<p>Accession number: PF01762 Definition: Galactosyltransferase Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_885 (release 4.2) Gathering cutoffs: -46 -46 Trusted cutoffs: -43.90 -43.90 Noise cutoffs: -49.80 -49.80 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98079080 Reference Title: Cloning of a human Reference Title: UDP-galactose:2-acetamido-2-deoxy-D-glucose 3beta- Reference Title: galactosyltransferase catalyzing the formation of type 1 Reference Title: chains. Reference Author: Kolbinger F, Streiff MB, Katopodis AG; Reference Location: J Biol Chem 1998;273:433-440. Reference Number: [2] Reference Medline: 98079027 Reference Title: Genomic cloning and expression of three</p>

Pfam	Prosite	Full Name	Description
			<p>murine</p> <p>Reference Title: UDP-galactose: beta-N- acetylglucosamine</p> <p>Reference Title: beta1,3-galactosyltransferase genes.</p> <p>Reference Author: Hennet T, Dinter A, Kuhnert P, Mattu TS, Rudd PM, Berger</p> <p>Reference Author: EG;</p> <p>Reference Location: J Biol Chem 1998;273:58-65.</p> <p>Database Reference: INTERPRO; IPR002659;</p> <p>Database reference: PFAMB; PB005938;</p> <p>Database reference: PFAMB; PB012965;</p> <p>Comment: This family includes the galactosyltransferases</p> <p>Comment: UDP-galactose:2-acetamido-2-deoxy-D-glucose3beta-galactosyltransferase</p> <p>Comment: Swiss:O43825 [1] and UDP-Gal:beta-GlcNAc beta 1,3-galactosyltransferase</p> <p>Comment: Swiss:O54904 [2].</p> <p>Comment: Specific galactosyltransferases transfer galactose to GlcNAc terminal</p> <p>Comment: chains in the synthesis of the lacto-series oligosaccharides types 1</p> <p>Comment: and 2 [1].</p> <p>Number of members: 29</p>
G-alpha		G-protein alpha subunit	<p>Accession number: PF00503</p> <p>Definition: G-protein alpha subunit</p> <p>Author: Finn RD</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_11 (release 1.0)</p> <p>Gathering cutoffs: 13.8 13.8</p> <p>Trusted cutoffs: 13.80 13.80</p> <p>Noise cutoffs: 9.70 12.70</p> <p>HMM build command line: hmmbuild -f HMM SEED</p> <p>HMM build command line: hmmbuild --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 94353239</p> <p>Reference Title: Structures of active conformations of Gi alpha 1 and the</p> <p>Reference Title: mechanism of GTP hydrolysis.</p> <p>Reference Author: Coleman DE, Berghuis AM, Lee E, Linder ME, Gilman AG,</p> <p>Reference Author: Sprang SR;</p> <p>Reference Location: Science 1994;265:1405-1412.</p> <p>Reference Number: [2]</p> <p>Reference Medline: 97004345</p> <p>Reference Title: How G proteins work: a continuing story.</p> <p>Reference Author: Coleman DE, Sprang SR;</p> <p>Reference Location: Trends Biochem Sci 1996;21:41-44.</p> <p>Database Reference: PRINTS; PR00318;</p> <p>Database Reference: SCOP; 1gia; fa; [SCOP-USA][CATH-PDBSUM]</p> <p>Database Reference: INTERPRO; IPR001019;</p> <p>Database Reference: PDB; 1gia ; 34; 343;</p> <p>Database Reference: PDB; 1gil ; 34; 343;</p> <p>Database Reference: PDB; 1as0 ; 32; 344;</p> <p>Database Reference: PDB; 1gfi ; 33; 345;</p> <p>Database Reference: PDB; 1as2 ; 32; 346;</p> <p>Database Reference: PDB; 1bh2 ; 32; 346;</p> <p>Database Reference: PDB; 1cip A; 32; 347;</p> <p>Database Reference: PDB; 1git ; 32; 348;</p> <p>Database Reference: PDB; 1agr D; 11; 353;</p> <p>Database Reference: PDB; 1gg2 A; 6; 348;</p> <p>Database Reference: PDB; 1gp2 A; 6; 348;</p> <p>Database Reference: PDB; 1bof ; 10; 353;</p> <p>Database Reference: PDB; 1as3 ; 9; 353;</p> <p>Database Reference: PDB; 1gdd ; 9; 353;</p> <p>Database Reference: PDB; 1agr A; 6; 353;</p> <p>Database Reference: PDB; 1tag ; 27; 340;</p> <p>Database Reference: PDB; 1tad A; 27; 342;</p> <p>Database Reference: PDB; 1tad B; 27; 342;</p> <p>Database Reference: PDB; 1tnd B; 27; 342;</p> <p>Database Reference: PDB; 1tnd C; 27; 342;</p> <p>Database Reference: PDB; 1tad C; 27; 344;</p> <p>Database Reference: PDB; 1tnd A; 27; 349;</p>

Pfam	Prosite	Full Name	Description
			<p>Database Reference PDB; 1cjk C; 39; 388;</p> <p>Database Reference PDB; 1cjt C; 39; 388;</p> <p>Database Reference PDB; 1cju C; 39; 388;</p> <p>Database Reference PDB; 1cqv C; 39; 388;</p> <p>Database Reference PDB; 1azt A; 35; 391;</p> <p>Database Reference PDB; 1azt B; 35; 391;</p> <p>Database Reference PDB; 1azs C; 36; 393;</p> <p>Database reference: PFAMB; PB034080;</p> <p>Comment: G proteins couple receptors of extracellular signals to intracellular</p> <p>Comment: signaling pathways.</p> <p>Comment: The G protein alpha subunit binds guanyl nucleotide and is a weak</p> <p>Comment: GTPase.</p> <p>Number of members: 245</p>
GCV_H		Glycine cleavage H-protein	<p>Accession number: PF01597</p> <p>Definition: Glycine cleavage H-protein</p> <p>Author: Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_988 (release 4.1)</p> <p>Gathering cutoffs: 25 25</p> <p>Trusted cutoffs: 27.90 27.90</p> <p>Noise cutoffs: -58.80 -58.80</p> <p>HMM build command line: hmmbuild -F HMM SEED</p> <p>HMM build command line: hmmscalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 94255425</p> <p>Reference Title: X-ray structure determination at 2.6-A resolution of a</p> <p>Reference Title: lipote- containing protein: the H-protein of the glycine</p> <p>Reference Title: decarboxylase complex from pea leaves.</p> <p>Reference Author: Pares S, Cohen-Addad C, Sieker L, Neuburger M, Douce R;</p> <p>Reference Location: Proc Natl Acad Sci U S A 1994;91:4850-4853.</p> <p>Database Reference: SCOP; 1htp; fa; [SCOP-USA][CATH-PDBSUM]</p> <p>Database Reference INTERPRO; IPR002930;</p> <p>Database Reference PDB; 1hpc A; 2; 127;</p> <p>Database Reference PDB; 1hpc B; 2; 127;</p> <p>Database Reference PDB; 1htp ; 2; 127;</p> <p>Comment: This is a family of glycine cleavage H-proteins, part of the glycine</p> <p>Comment: cleavage multienzyme complex (GCV)</p> <p>Comment: found in bacteria and the mitochondria</p> <p>Comment: of eukaryotes. GCV catalyses the catabolism of glycine in eukaryotes.</p> <p>Comment: A lipoyl group is attached to a completely conserved lysine residue.</p> <p>Comment: The H protein shuttles the methylamine group of glycine from the</p> <p>Comment: P protein to the T protein.</p> <p>Number of members: 40</p>
GCV_T		Glycine cleavage T-protein (aminomethyl transferase)	<p>Accession number: PF01571</p> <p>Definition: Glycine cleavage T-protein (aminomethyl transferase)</p> <p>Author: Bashton M, Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_933 (release 4.0)</p> <p>Gathering cutoffs: -146 -146</p> <p>Trusted cutoffs: -124.50 -124.50</p> <p>Noise cutoffs: -167.90 -167.90</p> <p>HMM build command line: hmmbuild -F HMM SEED</p> <p>HMM build command line: hmmscalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 97199363</p> <p>Reference Title: Cloning, and molecular characterization of the GCV1 gene</p> <p>Reference Title: encoding the glycine cleavage T-protein from Saccharomyces</p> <p>Reference Title: cerevisiae.</p>

09689930 "101300

Pfam	Prosite	Full Name	Description
			<p>Reference Author: McNeil JB, Zhang F, Taylor BV, Sinclair DA, Pearlman RE, Reference Author: Bogнар AL; Reference Location: Gene 1997;186:13-20. Database Reference INTERPRO; IPR002536; Database reference: PFAMB; PB004229; Comment: This is a family of glycine cleavage T-proteins, part of the glycine Comment: cleavage multienzyme complex (GCV) found in bacteria and the mitochondria Comment: of eukaryotes. GCV catalyses the catabolism of glycine in eukaryotes. Comment: The T-protein is an aminomethyl transferase. Number of members: 27</p>
G-gamma	PDOC01002	G-protein gamma subunit profile	<p>Guanine nucleotide-binding proteins (G proteins) [1] act as intermediaries in the transduction of signals generated by transmembrane receptors. G proteins consist of three subunits (alpha, beta, and gamma). The alpha subunit binds to and hydrolyzes GTP; the functions of the beta and gamma subunits are less clear but they seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition.</p> <p>The gamma subunits are small proteins (from 70 to 110 residues) that are bound to the membrane via a isoprenyl group (either a farnesyl or a geranyl-geranyl) covalently linked to their C-terminus. In mammals there are at least 12 different isoforms of gamma subunits.</p> <p>The <i>Caenorhabditis elegans</i> protein egl-10, which is a regulator of G-protein signalling, contains a G-protein gamma-like domain.</p> <p>We have developed a profile that spans the complete length of the gamma subunit.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Sequences known to belong to this class detected by the profile ALL, except for yeast and squid G-protein gamma. Other sequence(s) detected in SWISS-PROT NONE. Expert(s) to contact by email Pennington S.R. srpenn@liverpool.ac.uk</p> <p>Last update November 1997 / First entry. References [1] Pennington S.R. Protein Prof. 2:16-315(1995).</p>
glutaredoxin	PDOC00173	Glutaredoxin	<p>Glutaredoxin [1,2,3], also known as thioltransferase, is a small protein of approximately one hundred amino-acid residues. It functions as an electron carrier in the glutathione-dependent synthesis of deoxyribonucleotides by the enzyme ribonucleotide reductase. Like thioredoxin, which functions in a similar way, glutaredoxin possesses an active center disulfide bond. It exists in either a reduced or an oxidized form where the two cysteine residues are linked in an intramolecular disulfide bond.</p>

Pfam	Prosite	Full Name	Description
			<p>Glutaredoxin has been sequenced in a variety of species. On the basis of extensive sequence similarity, it has been proposed [4] that vaccinia protein O2L is most probably a glutaredoxin. Finally, it must be noted that phage T4 thioredoxin seems also to be evolutionary related.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [LIVD]-[FYSA]-x(4)-C-[PV]-[FYWH]-C-x(2)-[TAV]-x(2,3)-[LIV] [The two C's form the redox-active bond] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Note in position 5 of the pattern, all glutaredoxin sequences have Pro while T4 thioredoxin has Val. Last update December 1999 / Pattern and text revised. References [1] Gleason F.K., Holmgren A. FEMS Microbiol. Rev. 54:271-298(1988).</p> <p>[2] Holmgren A. Biochem. Soc. Trans. 16:95-96(1988).</p> <p>[3] Holmgren A. J. Biol. Chem. 264:13963-13966(1989).</p> <p>[4] Johnson G.P., Goebel S.J., Perkus M.E., Davis S.W., Winslow J.P., Paoletti E. Virology 181:378-381(1991).</p>
Glyco_hydro_1	PDOC00495	Glycosyl hydrolases family 1 signatures	<p>It has been shown [1 to 4] that the following glycosyl hydrolases can be, on the basis of sequence similarities, classified into a single family:</p> <ul style="list-style-type: none"> - Beta-glucosidases (EC 3.2.1.21) from various bacteria such as <i>Agrobacterium</i> strain ATCC 21400, <i>Bacillus polymyxa</i>, and <i>Caldocellum saccharolyticum</i>. - Two plants (clover) beta-glucosidases (EC 3.2.1.21). - Two different beta-galactosidases (EC 3.2.1.23) from the archaeobacteria <i>Sulfolobus solfataricus</i> (genes <i>bgaS</i> and <i>lacS</i>). - 6-phospho-beta-galactosidases (EC 3.2.1.85) from various bacteria such as <i>Lactobacillus casei</i>, <i>Lactococcus lactis</i>, and <i>Staphylococcus aureus</i>. - 6-phospho-beta-glucosidases (EC 3.2.1.86) from <i>Escherichia coli</i> (genes <i>bglB</i> and <i>ascB</i>) and from <i>Erwinia chrysanthemi</i> (gene <i>arbB</i>). - Plants myrosinases (EC 3.2.3.1) (sinigrinase) (thioglucosidase). - Mammalian lactase-phlorizin hydrolase (LPH) (EC 3.2.1.108 / EC 3.2.1.62). <p>LPH, an integral membrane glycoprotein, is the enzyme that splits lactose in the small intestine. LPH is a large protein of about 1900 residues which contains four tandem repeats of a domain of about 450 residues which is evolutionary related to the above glycosyl hydrolases.</p> <p>One of the conserved regions in these enzymes is centered on a conserved glutamic acid residue which has been shown [5], in the beta-</p>

Pfam	Prosite	Full Name	Description
			<p>glucosidase from Agrobacterium, to be directly involved in glycosidic bond cleavage by acting as a nucleophile. We have used this region as a signature pattern. As a second signature pattern we selected a conserved region, found in the N-terminal extremity of these enzymes, this region also contains a glutamic acid residue.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [LIVMFSTC]-[LIVFYS]-[LIV]-[LIVMST]-E-N-G-[LIVMFAR]-[CSAGN] [E is the active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 12.</p> <p>Note this pattern will pick up the last two domains of LPH; the first two domains, which are removed from the LPH precursor by proteolytic processing, have lost the active site glutamate and may therefore be inactive [4].</p> <p>Consensus pattern F-x-[FYWM]-[GSTA]-x-[GSTA]-x-[GSTA](2)-[FYNH]-[NQ]-x-E-x- [GSTA] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Note this pattern will pick up the last three domains of LPH. Expert(s) to contact by email Henrissat B. bernie@afmb.cnrs-mrs.fr</p> <p>Last update November 1995 / Patterns and text revised. References [1] Henrissat B. Biochem. J. 280:309-316(1991).</p> <p>[2] Henrissat B. Protein Seq. Data Anal. 4:61-62(1991).</p> <p>[3] Gonzalez-Candelas L., Ramon D., Polaina J. Gene 95:31-38(1990).</p> <p>[4] El Hassouni M., Henrissat B., Chippaux M., Barras F. J. Bacteriol. 174:765-777(1992).</p> <p>[5] Withers S.G., Warren R.A.J., Street I.P., Rupitz K., Kempton J.B., Aebersold R. J. Am. Chem. Soc. 112:5887-5889(1990).</p>
Glyco_hydro_19	PDOC00620	Chitinases family 19 signatures	<p>Chitinases (EC 3.2.1.14) [1] are enzymes that catalyze the hydrolysis of the beta-1,4-N-acetyl-D-glucosamine linkages in chitin polymers. From the view point of sequence similarity chitinases belong to either family 18 or 19 in the classification of glycosyl hydrolases [2,E1]. Chitinases of family 19 (also known as classes IA or I and IB or II) are enzymes from plants that function in the defense against fungal and insect pathogens by destroying their chitin-containing cell wall. Class IA/I and IB/II enzymes differ in the</p>

Pfam	Prosite	Full Name	Description
			<p>presence (IA/I) or absence (IB/II) of a N-terminal chitin-binding domain (see the relevant entry <PDOC00025>). The catalytic domain of these enzymes consist of about 220 to 230 amino acid residues.</p> <p>As signature patterns we selected two highly conserved regions, the first one is located in the N-terminal section and contains one of the six cysteines which are conserved in most, if not all, of these chitinases and which is probably involved in a disulfide bond.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern C-x(4,5)-F-Y-[ST]-x(3)-[FY]-[LIVMF]-x-A-x(3)-[YF]-x(2)-F- [GSA] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Consensus pattern [LIVM]-[GSA]-F-x-[STAG](2)-[LIVMFY]-W-[FY]-W-[LIVM] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Expert(s) to contact by email Neuhaus J.-M. jean-marc.neuhaus@bota.unine.ch</p> <p>Henrissat B. bernie@afmb.cnrs-mrs.fr</p> <p>Last update November 1997 / Text revised. References [1] Flach J., Pilet P.-E., Jolles P. Experientia 48:701-716(1992).</p> <p>[2] Henrissat B. Biochem. J. 280:309-316(1991).</p> <p>[E1] http://www.expasy.ch/cgi-bin/lists?glycosid.txt</p>
Glyco_hydro_3_C	PDOC00621	Glycosyl hydrolases family 3 active site	<p>It has been shown [1,2] that the following glycosyl hydrolases can be, on the basis of sequence similarities, classified into a single family:</p> <ul style="list-style-type: none"> - Beta glucosidases (EC 3.2.1.21) from the fungi <i>Aspergillus wentii</i> (A-3), <i>Hansenula anomala</i>, <i>Kluyveromyces fragilis</i>, <i>Saccharomycopsis fibuligera</i>, (BGL1 and BGL2), <i>Schizophyllum commune</i> and <i>Trichoderma reesei</i> (BGL1). - Beta glucosidases from the bacteria <i>Agrobacterium tumefaciens</i> (Cbg1), <i>Butyrivibrio fibrisolvens</i> (bglA), <i>Clostridium thermocellum</i> (bglB), <i>Escherichia coli</i> (bglX), <i>Erwinia chrysanthemi</i> (bgxA) and <i>Ruminococcus albus</i>. - <i>Alteromonas</i> strain O-7 beta-hexosaminidase A (EC 3.2.1.52). - <i>Bacillus subtilis</i> hypothetical protein yzbA. - <i>Escherichia coli</i> hypothetical protein ycfO and HI0959, the corresponding <i>Haemophilus influenzae</i> protein. <p>One of the conserved regions in these enzymes is centered on a conserved aspartic acid residue which has been shown [3], in <i>Aspergillus</i></p>

00669930 101300

Pfam	Prosite	Full Name	Description
			<p>wentii beta-glucosidase A3, to be implicated in the catalytic mechanism. We have used this region as a signature pattern.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [LIVM](2)-[KR]-x-[EQK]-x(4)-G-[LIVMFT]-[LIVT]-[LIVMF]-[ST]-D-x(2)-[SGADNI] [D is the active site residue]</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Expert(s) to contact by email Henrissat B. bernie@afmb.cnrs-mrs.fr</p> <p>Last update November 1997 / Pattern and text revised.</p> <p>References [1] Henrissat B. Biochem. J. 280:309-316(1991).</p> <p>[2] Castle L.A., Smith K.D., Morris R.O. J. Bacteriol. 174:1478-1486(1992).</p> <p>[3] Bause E., Legler G. Biochim. Biophys. Acta 626:459-465(1980).</p>
Glyco_hydro_45	PDOC00877	Glycosyl hydrolases family 45 active site	<p>The microbial degradation of cellulose and xylans requires several types of enzymes such as endoglucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91) (exoglucanases), or xylanases (EC 3.2.1.8) [1,2]. Fungi and bacteria produces a spectrum of cellulolytic enzymes (cellulases) and xylanases which, on the basis of sequence similarities, can be classified into families. One of these families is known as the cellulase family K or as the glycosyl hydrolases family 45 [3,E1]. The enzymes which are currently known to belong to this family are listed below.</p> <ul style="list-style-type: none"> - Endoglucanase 5 from Humicola insolens. - Endoglucanase 5 from Trichoderma reesei (egl5). - Endoglucanase K from Fusarium oxysporum. - Endoglucanase B from Pseudomonas fluorescens (celB). - Endoglucanase 1 from Ustilago maydis (egl1). <p>The best conserved regions in these enzymes is located in the N-terminal section. It contains an aspartic acid residue which has been shown [4] to act as a nucleophile in the catalytic mechanism. We use this region as a signature pattern.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [STA]-T-R-Y-[FYW]-D-x(5)-[CA] [The D is an active site residue]</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Expert(s) to contact by email Henrissat B. bernie@afmb.cnrs-mrs.fr</p>

Pfam	Prosite	Full Name	Description
			<p>Last update November 1997 / Pattern and text revised.</p> <p>References</p> <p>[1] Beguin P. Annu. Rev. Microbiol. 44:219-248(1990).</p> <p>[2] Gilkes N.R., Henrissat B., Kilburn D.G., Miller R.C. Jr., Warren R.A.J. Microbiol. Rev. 55:303-315(1991).</p> <p>[3] Henrissat B., Bairoch A. Biochem. J. 293:781-788(1993).</p> <p>[4] Davies G.J., Dodson G.G., Hubbard R.E., Tolley S.P., Dauter Z., Wilson K.S., Hjort C., Mikkelsen J.M., Rasmussen G., Schuelein M. Nature 365:362-364(1993).</p> <p>[E1] http://www.expasy.ch/cgi-bin/lists?glycosid.txt</p>
Glyco_hydro_47		Glycosyl hydrolase family 47	Members of this family are alpha-mannosidases that catalyse the hydrolysis of the terminal 1,2-linked alpha-D-mannose residues in the oligo-mannose oligosaccharide Man(9)(GlcNAc)(2). These enzymes are capable of taking part in the glycosylation pathway and glycoprotein processing.
GTP_cyclohydrol	PDOC00672	GTP cyclohydrolase I signatures	<p>GTP cyclohydrolase I (EC 3.5.4.16) catalyzes the biosynthesis of formic acid and dihydroneopterin triphosphate from GTP. This reaction is the first step in the biosynthesis of tetrahydrofolate in prokaryotes, of tetrahydrobiopterin in vertebrates, and of pteridine-containing pigments in insects.</p> <p>GTP cyclohydrolase I is a protein of from 190 to 250 amino acid residues. The comparison of the sequence of the enzyme from bacterial and eukaryotic sources shows that the structure of this enzyme has been extremely well conserved throughout evolution [1].</p> <p>As signature patterns we selected two conserved regions. The first contains a perfectly conserved tetrapeptide which is part of the GTP-binding pocket [2], the second region also contains conserved residues involved in GTP-binding.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [DEN]-[LIVM](2)-x(2)-[KRNQ]-[DEN]-[LIVM]-x(3)-[ST]-x-C-E- H-H Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Consensus pattern [SA]-x-[RK]-x-Q-[LIVM]-Q-E-[RN]-[LI]-[TSN] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update July 1999 / Patterns and text revised.</p> <p>References</p> <p>[1] Maier J., Witter K., Guetlich M., Ziegler I., Werner T., Ninnemann H.</p>

Pfam	Prosite	Full Name	Description
			Biochem. Biophys. Res. Commun. 212:705-711(1995). [2] Nar H., Huber R., Meining W., Schmid C., Weinkauff S., Bacher A. Structure 3:459-466(1995).
HCV_capsid		Hepatitis C virus capsid protein	Family members include nucleocapsid proteins of the HCV. This virus family comprises a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein c and mRNA. Uses for these polypeptides include: immunological epitopes for vaccines; or as mRNA chaperone proteins to aid in processing or to prevent degradation. References describing examples of these capsid polypeptides include: Chen et al., Virology 188:102-113(1992); and Okamoto et al., J. Gen. Virol. 72:2697-2704(1991
HD		HD domain	Accession number: PF01966 Definition: HD domain Author: Enright A, Ouzounis C, Bateman A Alignment method of seed: Clustalw Source of seed members: Enright A Gathering cutoffs: -1 -1 Trusted cutoffs: -0.50 -0.50 Noise cutoffs: -2.50 -2.50 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 99085258 Reference Title: The HD domain defines a new superfamily of metal-dependent Reference Title: phosphohydrolases. Reference Author: Aravind L, Koonin EV; Reference Location: Trends Biochem Sci 1998;23:469-472. Database Reference: INTERPRO; IPR002819; Database reference: PFAMB; PB005654; Database reference: PFAMB; PB006725; Database reference: PFAMB; PB009617; Database reference: PFAMB; PB012663; Database reference: PFAMB; PB035384; Database reference: PFAMB; PB040597; Comment: HD domains are metal dependent phosphohydrolases. Number of members: 63
HDV_ag		Hepatitis delta virus delta antigen	Accession number: PF01517 Definition: Hepatitis delta virus delta antigen Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_808 (release 4.0) Gathering cutoffs: -8 -8 Trusted cutoffs: 23.30 23.30 Noise cutoffs: -40.50 -40.50 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 94065676 Reference Title: Characterization of RNA-binding domains of hepatitis delta Reference Title: antigen. Reference Author: Poisson F, Roingeard P, Baillou A, Dubois F, Bonelli F, Reference Author: Calogero RA, Goudeau A; Reference Location: J Gen Virol 1993;74:2473-2478. Reference Number: [2] Reference Medline: 98362586 Reference Title: Structural basis of the oligomerization of hepatitis delta Reference Title: antigen. Reference Author: Zuccola HJ, Rozzelle JE, Lemon SM, Erickson BW, Hogle JM; Reference Location: Structure 1998;6:821-830. Database Reference: SCOP; 1a92; fa; [SCOP-USA][CATH-PDBSUM]

Pfam	Prosite	Full Name	Description
			<p>Database Reference INTERPRO; IPR002506;</p> <p>Database Reference PDB; 1a92 A; 12; 23;</p> <p>Database Reference PDB; 1a92 B; 12; 23;</p> <p>Database Reference PDB; 1a92 C; 12; 23;</p> <p>Database Reference PDB; 1a92 D; 12; 60;</p> <p>Database Reference PDB; 1a92 A; 47; 60;</p> <p>Database Reference PDB; 1a92 B; 47; 60;</p> <p>Database Reference PDB; 1a92 C; 47; 60;</p> <p>Comment: The hepatitis delta virus (HDV) encodes a single protein, the</p> <p>Comment: hepatitis delta antigen (HDAg). The central region of this protein</p> <p>Comment: has been shown to bind RNA [1]. Several interactions are also</p> <p>Comment: mediated by a coiled-coil region at the N terminus of the protein [2].</p> <p>Number of members: 145</p>
hemolysinCabind	PD0C00293	Hemolysin-type calcium-binding region signature	<p>Gram-negative bacteria produce a number of proteins which are secreted into the growth medium by a mechanism that does not require a cleaved N-terminal signal sequence. These proteins, while having different functions, seem [1] to share two properties: they bind calcium and they contain a variable number of tandem repeats consisting of a nine amino acid motif rich in glycine, aspartic acid and asparagine. It has been shown [2] that such a domain is involved in the binding of calcium ions in a parallel beta roll structure. The proteins which are currently known to belong to this category are:</p> <ul style="list-style-type: none"> - Hemolysins from various species of bacteria. Bacterial hemolysins are exotoxins that attack blood cell membranes and cause cell rupture. The hemolysins which are known to contain such a domain are those from: <i>E. coli</i> (gene hlyA), <i>A. pleuropneumoniae</i> (gene appA), <i>A. actinomycetemcomitans</i> and <i>P. haemolytica</i> (leukotoxin) (gene lktA). - Cyclolysin from <i>Bordetella pertussis</i> (gene cyaA). A multifunctional protein which is both an adenylate cyclase and a hemolysin. - Extracellular zinc proteases: serralysin (EC 3.4.24.40) from <i>Serratia</i>, prtB and prtC from <i>Erwinia chrysanthemi</i> and aprA from <i>Pseudomonas aeruginosa</i>. - Nodulation protein nodO from <i>Rhizobium leguminosarum</i>. <p>We derived a signature pattern from conserved positions in the sequence of the calcium-binding domain.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern D-x-[LI]-x(4)-G-x-D-x-[LI]-x-G-G-x(3)-D</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Note this pattern is found once in nodO and the extracellular proteases but up to 5 times in some hemolysin/cyclolysins.</p> <p>Last update October 1993 / Text revised.</p> <p>References [1] Economou A., Hamilton W.D.O., Johnston A.W.B., Downie J.A. EMBO J. 9:349-354(1990).</p>

Pfam	Prosite	Full Name	Description
			[2] Baumann U., Wu S., Flaherty K.M., McKay D.B. EMBO J. 12:3357-3364(1993).
Heptosyltranf		Heptosyltransferase	Accession number: PF01075 Definition: Heptosyltransferase Author: Finn RD, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_839 (release 3.0) Gathering cutoffs: -40 -40 Trusted cutoffs: -31.80 -31.80 Noise cutoffs: -47.10 -47.10 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98112827 Reference Title: Enzymatic synthesis of lipopolysaccharide in Escherichia Reference Title: coli. Purification and properties of heptosyltransferase I. Reference Author: Kadrmas JL, Raetz CR; Reference Location: J Biol Chem 1998;273:2799-2807. Database Reference: INTERPRO; IPR002201; Database reference: PFAMB; PB021100; Database reference: PFAMB; PB033445; Database reference: PFAMB; PB041423; Comment: Lipopolysaccharide is a major component of the outer leaflet of Comment: the outer membrane in Gram-negative bacteria. It is composed of Comment: three domains; lipid A, Core oligosaccharide and the O-antigen. Comment: All of these enzymes transfer heptose to the lipopolysaccharide Comment: core. Number of members: 46
Herpes_alk_exo		Herpesvirus alkaline exonuclease	Accession number: PF01771 Definition: Herpesvirus alkaline exonuclease Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_822 (release 4.2) Gathering cutoffs: 25 25 Trusted cutoffs: 318.00 318.00 Noise cutoffs: -277.60 -277.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 85107093 Reference Title: Studies on the herpes simplex virus alkaline nuclease: Reference Title: detection of type-common and type-specific epitopes on the Reference Title: enzyme. Reference Author: Banks LM, Halliburton IW, Purifoy DJ, Killington RA, Powell Reference Author: KL; Reference Location: J Gen Virol 1985;66:1-14. Database Reference: INTERPRO; IPR001616; Comment: This family includes various alkaline exonucleases from Comment: members of the herpesviridae. Alkaline exonuclease Comment: appears to have an important role in the replication of Comment: herpes simplex virus [1]. Number of members: 23
Herpes_gl		Alpha herpesvirus glycoprotein I	Accession number: PF01688 Definition: Alpha herpesvirus glycoprotein I Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1222 (release 4.1) Gathering cutoffs: 25 25

Pfam	Prosite	Full Name	Description
			<p>Trusted cutoffs: 157.20 157.20 Noise cutoffs: -126.70 -126.70 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild --seed 0 HMM Reference Number: [1] Reference Medline: 96357074 Reference Title: Biosynthesis of glycoproteins E and I of feline Reference Title: herpesvirus: gE-gI interaction is required for intracellular transport. Reference Author: Mijnes JD, van der Horst LM, van Anken E, Horzinek MC, Reference Author: Rottier PJ, de Groot RJ; Reference Location: J Virol 1996;70:5466-5475. Reference Number: [2] Reference Medline: 94267406 Reference Title: Identification of the feline herpesvirus type 1 (FHV-1) Reference Title: genes encoding glycoproteins G, D, I and E: expression of Reference Title: FHV-1 glycoprotein D in vaccinia and raccoon poxviruses. Reference Author: Spatz SJ, Rota PA, Maes RK; Reference Location: J Gen Virol 1994;75:1235-1244. Reference Number: [3] Reference Medline: 94267879 Reference Title: Unusual phosphorylation sequence in the gpIV (gI) component Reference Title: of the varicella-zoster virus gpI-gpIV glycoprotein complex Reference Title: (VZV gE-gI complex). Reference Author: Yao Z, Grose C; Reference Location: J Virol 1994;68:4204-4211. Database Reference: INTERPRO; IPR002874; Comment: This family consists of glycoprotein I form various members of the Comment: alphaherpesvirinae these include herpesvirus, varicella-zoster virus Comment: and pseudorabies virus. Glycoprotein I (gI) is important during natural Comment: infection, mutants lacking gI produce smaller lesions at the site of Comment: infection and show reduced neuronal spread [1]. gI forms a heterodimeric Comment: complex with gE; this complex displays Fc receptor activity (binds to Comment: the Fc region of immunoglobulin) [1]. Glycoproteins are also important Comment: in the production of virus-neutralizing antibodies and cell mediated Comment: immunity [2]. The alphaherpesvirinae have a dsDNA genome and have no Comment: RNA stage during viral replication. Number of members: 22</p>
Herpes_glycop_D		Herpesvirus glycoprotein M	<p>Accession number: PF01528 Definition: Herpesvirus glycoprotein M Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_929 (release 4.0) Gathering cutoffs: 25 25 Trusted cutoffs: 197.30 197.30 Noise cutoffs: -229.70 -229.70 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild --seed 0 HMM Reference Number: [1] Reference Medline: 96357105 Reference Title: Identification and characterization of pseudorabies virus Reference Title: glycoprotein gM as a nonessential virion component. Reference Author: Dijkstra JM, Visser N, Mettenleiter TC, Klupp BG; Reference Location: J Virol 1996;70:5684-5688.</p>

Pfam	Prosite	Full Name	Description
			<p>Reference Number: [2] Reference Medline: 95381611 Reference Title: Identification and molecular characterization of the murine Reference Title: cytomegalovirus homolog of the human cytomegalovirus UL100 Reference Title: gene. Reference Author: Li W, Eidman K, Gehrz RC, Kari B; Reference Location: Virus Res 1995;36:163-175. Database Reference INTERPRO; IPR000785; Comment: The herpesvirus glycoprotein M (gM) is an integral membrane protein Comment: predicted to contain 8 transmembrane segments [2]. Glycoprotein M is Comment: not essential for viral replication [1]. Number of members: 24</p>
HesB-like	PDOC00887	Hypothetical hesB/yadR/yfhF family signature	<p>The following uncharacterized proteins have been shown [1] to share regions of similarities:</p> <ul style="list-style-type: none"> - Anabaena and related cyanobacteria protein hesB which may be required for nitrogen fixation. - Escherichia coli hypothetical protein yadR and HI1723, the corresponding Haemophilus influenzae protein. - Escherichia coli hypothetical protein ydiC. - Escherichia coli hypothetical protein yfhF and HI0376, the corresponding Haemophilus influenzae protein. - Mycobacterium tuberculosis hypothetical protein Rv2204c. - Synechocystis strain PCC 6803 hypothetical protein slr1417. - Synechocystis strain PCC 6803 hypothetical protein slr1565. - A hypothetical protein in the nifU 5' region of many nitrogen fixing bacteria. - Porphyra purpurea chloroplast hypothetical protein in apcF-rps4 intergenic region. - Yeast hypothetical protein YLL027W. - Yeast hypothetical protein YPR067W. <p>These are small proteins (106 to 135 amino-acid residues in bacteria, about 200 residues in fungi) that contain a number of conserved regions. The most noteworthy of these regions is located in the C-terminal extremity, it contains two conserved cysteines. We have used this region as a signature pattern.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern F-x-[LIVMFY]-x-N-[PG]-[NSKQ]-x(4)-C-x-C-[GS]-x-S-F Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update December 1999 / Pattern and text revised. References [1] Bairoch A., Rudd K.E. Unpublished observations (1995).</p>
HisG	PDOC01020	ATP phosphoribosyltransferase signature	<p>ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern we</p>

Pfam	Prosite	Full Name	Description
			<p>(EC 1.1.1.88) to deacetylate mevalonate into 3-hydroxy-3-methylglutaryl-CoA [3]. The Pseudomonas enzyme is structurally related to the catalytic domain of NADP-dependent HMG-CoA reductases.</p> <p>We selected three conserved regions as signature patterns for HMG-CoA reductases. The first is located in the center of the catalytic domain, the second is a glycine-rich region located in the C-terminal section of the same catalytic domain and the third is also located in the C-terminal section and contains an histidine residue that seems [4] to be implicated in the catalytic mechanism as a general base.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [RKH]-x(6)-D-x-M-G-x-N-x-[LIVMA] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 4.</p> <p>Consensus pattern [LIVM]-G-x-[LIVM]-G-G-[AG]-T Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT 5.</p> <p>Consensus pattern A-[LIVM]-x-[STAN]-x(2)-[LI]-x-[KRNQ]-[GSA]-H-[LM]-x- [FYLH] [H is an active site residue] Sequences known to belong to this class detected by the pattern ALL, except for archaeobacterial HMG-CoA reductases. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so.</p> <p>Last update November 1997 / Patterns and text revised; profile added.</p> <p>References [1] Caelles C., Ferrer A., Balcells L., Hegardt F.G., Boronat A. Plant Mol. Biol. 13:627-638(1989).</p> <p>[2] Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J. Mol. Cell. Biol. 8:3797-3808(1988).</p> <p>[3] Lam W.L., Doolittle W.F. J. Biol. Chem. 267:5829-5834(1992).</p> <p>[4] Beach M.J., Rodwell V.W. J. Bacteriol. 171:2994-3001(1989).</p> <p>[5] Darnay B.G., Wang Y., Rodwell V.W. J. Biol. Chem. 267:15064-15070(1992).</p>
HMGL-like	PDOC00813 PDOC00643	Hydroxymethylglutaryl-coenzyme A lyase active site; Alpha-isopropylmalate and homocitrate	3-hydroxy-3-methylglutaryl-coenzyme A lyase (HMG-CoA lyase or HL) (EC 4.1.3.4) catalyzes the transformation of HMG-CoA into acetyl-CoA and acetoacetate. In vertebrates it is a mitochondrial enzyme which is involved in

Pfam	Prosite	Full Name	Description
			<p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update November 1997 / Patterns and text revised.</p> <p>References [1] Wang S.-Z., Dean D.R., Chen J.-S., Johnson J.L. J. Bacteriol. 173:3041-3046(1991).</p>
hormone5	PDOC00237	Neurohypophysial hormones signature	<p>Oxytocin (or ocytocin) and vasopressin [1] are small (nine amino acid residues), structurally and functionally related neurohypophysial peptide hormones. Oxytocin causes contraction of the smooth muscle of the uterus and of the mammary gland while vasopressin has a direct antidiuretic action on the kidney and also causes vasoconstriction of the peripheral vessels. Like the majority of active peptides, both hormones are synthesized as larger protein precursors that are enzymatically converted to their mature forms. Peptides belonging to this family are also found in birds, fish, reptiles and amphibians (mesotocin, isotocin, valitocin, glutitocin, aspartocin, vasotocin, seritocin, asvatocin, phasvatocin), in worms (annetocin), octopi (cephalotocin), locust (locupressin or neuropeptide F1/F2) and in molluscs (conopressins G and S) [2].</p> <p>The pattern developed to detect this category of peptides spans their entire sequence and includes four invariant amino acid residues.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern C-[LIFY](2)-x-N-[CS]-P-x-G [The two C's are linked by a disulfide bond]. Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update November 1995 / Pattern and text revised.</p> <p>References [1] Acher R., Chauvet J. Biochimie 70:1197-1207(1988).</p> <p>[2] Chauvet J., Michel G., Ouedraogo Y., Chou J., Chait B.T., Acher R. Int. J. Pept. Protein Res. 45:482-487(1995).</p>
HPPK	PDOC00631	7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase signature	<p>All organisms require reduced folate cofactors for the synthesis of a variety of metabolites. Most microorganisms must synthesize folate de novo because they lack the active transport system of higher vertebrate cells which allows these organisms to use dietary folates. Enzymes involved in folate biosynthesis are therefore targets for a variety of antimicrobial agents such as trimethoprim or sulfonamides.</p> <p>7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase (EC 2.7.6.3) (HPPK) catalyzes the attachment of pyrophosphate to 6-hydroxymethyl-7,8-dihydropterin to form 6-hydroxymethyl-7,8-dihydropteridine pyrophosphate.</p>

09689800 101300

Pfam	Prosite	Full Name	Description
			<p>This is the first step in a three-step pathway leading to 7,8-dihydrofolate.</p> <p>Bacterial HPPK (gene folK or sulD) [1] is a protein of 160 to 270 amino acids. In the lower eukaryote <i>Pneumocystis carinii</i>, HPPK is the central domain of a multifunctional folate synthesis enzyme (gene fas) [2].</p> <p>As a signature for HPPK, we selected a conserved region located in the central section of these enzymes.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [KRHD]-x-[GA]-[PSAE]-R-x(2)-D-[LIV]-D-[LIVM](2)</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update July 1999 / Pattern and text revised.</p> <p>References [1] Talarico T.L., Ray P.H., Dev I.K., Merrill B.M., Dallas W.S. J. Bacteriol. 174:5971-5977(1992).</p> <p>[2] Volpes F., Dyer M., Scaife J.G., Darby G., Stammers D.K., Delves C.J. Gene 112:213-218(1992).</p>
HTH_AraC	PDOC00040	Bacterial regulatory proteins, araC family signature and profile	<p>The many bacterial transcription regulation proteins which bind DNA through a 'helix-turn-helix' motif can be classified into subfamilies on the basis of sequence similarities. One of these subfamilies groups together the following proteins [1,2]:</p> <ul style="list-style-type: none"> - aarP, a transcriptional activator of the 2'-N-acetyltransferase gene in <i>Providencia stuartii</i>. - ada, an <i>Escherichia coli</i> and <i>Salmonella typhimurium</i> bifunctional protein that repairs alkylated guanine in DNA by transferring the alkyl group at the O(6) position to a cysteine residue in the enzyme. The methylated protein acts a positive regulator of its own synthesis and of the alkA, alkB and aidB genes. - adaA, a <i>Bacillus subtilis</i> bifunctional protein that acts both as a transcriptional activator of the ada operon and as a methylphosphotriester-DNA alkyltransferase. - adiY, an <i>Escherichia coli</i> protein of unknown function. - aggR, the transcriptional activator of aggregative adherence fimbria I expression in enteroaggregative <i>Escherichia coli</i>. - appY, a protein which acts as a transcriptional activator of acid phosphatase and other proteins during the deceleration phase of growth and acts as a repressor for other proteins that are synthesized in exponential growth or in the stationary phase. - araC, the arabinose operon regulatory protein, which activates the transcription of the araBAD genes. - cafR, the <i>Yersinia pestis</i> F1 operon positive regulatory protein.

09669960 "00EFO"

Pfam	Prosite	Full Name	Description
			<ul style="list-style-type: none"> - celD, the Escherichia coli cel operon repressor. - cfaD, a protein which is required for the expression of the CFA/I adhesin of enterotoxigenic Escherichia coli. - csvR, a transcriptional activator of fimbrial genes in enterotoxigenic Escherichia coli. - envY, the porin thermoregulatory protein, which is involved in the control of the temperature-dependent expression of several Escherichia coli envelope proteins such as ompF, ompC, and lamB. - exsA, an activator of exoenzyme S synthesis in Pseudomonas aeruginosa. - fapR, the positive activator for the expression of the 987P operon coding for the fimbrial protein in enterotoxigenic Escherichia coli. - hrpB, a positive regulator of pathogenicity genes in Burkholderia solanacearum. - invF, the Salmonella typhimurium invasion operon regulator. - marA, which may be a transcriptional activator of genes involved in the multiple antibiotic resistance (mar) phenotype. - melR, the melibiose operon regulatory protein, which activates the transcription of the melAB genes. - mixE, a Shigella flexneri protein necessary for secretion of ipa invasins. - mmsR, the transcriptional activator for the mmsAB operon in Pseudomonas aeruginosa. - msmR, the multiple sugar metabolism operon transcriptional activator in Streptococcus mutans. - pchR, a Pseudomonas aeruginosa activator for pyochelin and ferripyochelin receptor. - perA, a transcriptional activator of the eaeA gene for intimin in enteropathogenic Escherichia coli. - pocR, a Salmonella typhimurium regulator of the cobalamin biosynthesis operon. - pqrA, from Proteus vulgaris. - rafR, the regulator of the raffinose operon in Pedicoccus pentosaceus. - ramA, from Klebsiella pneumoniae. - rhaR, the Escherichia coli and Salmonella typhimurium L-rhamnose operon transcriptional activator. - rhaS, an Escherichia coli and Salmonella typhimurium positive activator of genes required for rhamnose utilization. - rns, a protein which is required for the expression of the cs1 and cs2 adhesins of enterotoxigenic Escherichia coli. - rob, a protein which binds to the right arm of the replication origin oriC of the Escherichia coli chromosome. - soxS, a protein that, with the soxR protein, controls a superoxide response regulon in Escherichia coli. - tetD, a protein from transposon TN10. - tcpN or toxT, the Vibrio cholerae transcriptional activator of the tcp operon involved in pilus biosynthesis and transport. - thcR, a probable regulator of the thc operon for the degradation of the thiocarbamate herbicide EPTC in Rhodococcus sp. strain N186/21. - ureR, the transcriptional activator of the plasmid-encoded urease operon in Enterobacteriaceae.

09669980 "09669980"

Pfam	Prosite	Full Name	Description
			<p>Nucleic Acids Res. 21:807-810(1993).</p> <p>[2] Henikoff S., Wallace J.C., Brown J.P. Meth. Enzymol. 183:111-132(1990).</p> <p>[3] Bustos S.A., Schleif R.F. Proc. Natl. Acad. Sci. U.S.A. 90:5638-5642(1993).</p>
Hydrolase		haloacid dehalogenase-like hydrolase	<p>Accession number: PF00702 Definition: haloacid dehalogenase-like hydrolase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_566 (release 2.1) Gathering cutoffs: 7 7 Trusted cutoffs: 7.10 7.10 Noise cutoffs: 2.90 2.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96355356 Reference Title: Crystal structure of L-2-haloacid dehalogenase from Reference Title: Pseudomonas sp. YL. An alpha/beta hydrolase structure that Reference Title: is different from the alpha/beta hydrolase fold. Reference Author: Hisano T, Hata Y, Fujii T, Liu JQ, Kurihara T, Esaki N, Reference Author: Soda K; Reference Location: J Biol Chem 1996;271:20322-20330. Database Reference: SCOP; 1jud; sf; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR001454; Database Reference PDB; 1jud ; 4; 197; Database Reference PDB; 1zrm ; 4; 197; Database Reference PDB; 1zrn ; 4; 197; Database Reference PDB; 1aq6 A; 2; 193; Database Reference PDB; 1aq6 B; 2; 193; Database Reference PDB; 1qq5 A; 2; 193; Database Reference PDB; 1qq5 B; 2; 193; Database Reference PDB; 1qq6 A; 2; 193; Database Reference PDB; 1qq6 B; 2; 193; Database Reference PDB; 1qq7 A; 2; 193; Database Reference PDB; 1qq7 B; 2; 193; Database Reference PDB; 1cqz A; 4; 19; Database Reference PDB; 1cr6 A; 4; 19; Database Reference PDB; 1cqz B; 4; 206; Database Reference PDB; 1cr6 B; 4; 206; Database Reference PDB; 1cqz A; 48; 206; Database Reference PDB; 1cr6 A; 48; 206; Database reference: PFAMB; PB000701; Database reference: PFAMB; PB001048; Database reference: PFAMB; PB019234; Database reference: PFAMB; PB032787; Database reference: PFAMB; PB040985; Database reference: PFAMB; PB041061; Database reference: PFAMB; PB041182; Database reference: PFAMB; PB041477; Database reference: PFAMB; PB041535; Database reference: PFAMB; PB041628; Database reference: PFAMB; PB041677; Comment: This family are structurally different from the alpha/ Comment: beta hydrolase family (abhydrolase). Comment: This family includes L-2-haloacid dehalogenase, epoxide Comment: hydrolases and phosphatases. Comment: The structure of the family consists of two domains. One Comment: is an inserted four helix bundle, which is the least well</p>

00669980 " 00669980

Pfam	Prosite	Full Name	Description
			<p>Reference Location: Protein Sci 1995;4:1608-1617.</p> <p>Comment: This domain of unknown function is found at the C-terminus</p> <p>Comment: of several transcription initiation factors [1].</p> <p>Number of members: 31</p>
ig	PDOC00262	Immunoglobulins and major histocompatibility complex proteins signature	<p>The basic structure of immunoglobulin (Ig) [1] molecules is a tetramer of two light chains and two heavy chains linked by disulfide bonds. There are two types of light chains: kappa and lambda, each composed of a constant domain (CL) and a variable domain (VL). There are five types of heavy chains: alpha, delta, epsilon, gamma and mu, all consisting of a variable domain (VH) and three (in alpha, delta and gamma) or four (in epsilon and mu) constant domains (CH1 to CH4).</p> <p>The major histocompatibility complex (MHC) molecules are made of two chains. In class I [2] the alpha chain is composed of three extracellular domains, a transmembrane region and a cytoplasmic tail. The beta chain (beta-2-microglobulin) is composed of a single extracellular domain. In class II [3], both the alpha and the beta chains are composed of two extracellular domains, a transmembrane region and a cytoplasmic tail.</p> <p>It is known [4,5] that the Ig constant chain domains and a single extracellular domain in each type of MHC chains are related. These homologous domains are approximately one hundred amino acids long and include a conserved intradomain disulfide bond. We developed a small pattern around the C-terminal cysteine involved in this disulfide bond which can be used to detect these category of Ig related proteins.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [FY]-x-C-x-[VA]-x-H-Sequences known to belong to this class detected by the pattern: Ig heavy chains type Alpha C region : All, in CH2 and CH3. Ig heavy chains type Delta C region : All, in CH3. Ig heavy chains type Epsilon C region : All, in CH1, CH3 and CH4. Ig heavy chains type Gamma C region : All, in CH3 and also CH1 in some cases Ig heavy chains type Mu C region : All, in CH2, CH3 and CH4. Ig light chains type Kappa C region : In all CL except rabbit and Xenopus. Ig light chains type Lambda C region : In all CL except rabbit. MHC class I alpha chains : All, in alpha-3 domains, including in the cytomegalovirus MHC-1 homologous protein [6]. Beta-2-microglobulin : All. MHC class II alpha chains: All, in alpha-2 domains. MHC class II beta chains: All, in beta-2 domains.</p> <p>Other sequence(s) detected in SWISS-PROT 71.</p> <p>Last update May 1991 / Text revised.</p> <p>References [1] Gough N. Trends Biochem. Sci. 6:203-205(1981).</p> <p>[2] Klein J., Figueroa F. Immunol. Today 7:41-44(1986).</p> <p>[3]</p>

00669980 " 101300

Pfam	Prosite	Full Name	Description
			<p>Figueroa F., Klein J. Immunol. Today 7:78-81(1986).</p> <p>[4] Orr H.T., Lancet D., Robb R.J., Lopez de Castro J.A., Strominger J.L. Nature 282:266-270(1979).</p> <p>[5] Cushley W., Owen M.J. Immunol. Today 4:88-92(1983).</p> <p>[6] Beck S., Barrel B.G. Nature 331:269-272(1988).</p>
IMPDH_C	PDOC00391	IMP dehydrogenase / GMP reductase signature	<p>IMP dehydrogenase (EC 1.1.1.205) (IMPDH) catalyzes the rate-limiting reaction of de novo GTP biosynthesis, the NAD-dependent reduction of IMP into XMP [1]. Inhibition of IMP dehydrogenase activity results in the cessation of DNA synthesis. As IMP dehydrogenase is associated with cell proliferation, it is a possible target for cancer chemotherapy. Mammalian and bacterial IMPDHs are tetramers of identical chains. There are two IMP dehydrogenase isozymes in humans [2].</p> <p>GMP reductase (EC 1.6.6.8) catalyzes the irreversible and NADPH-dependent reductive deamination of GMP into IMP [3]. It converts nucleobase, nucleoside and nucleotide derivatives of G to A nucleotides, and maintains intracellular balance of A and G nucleotides.</p> <p>IMP dehydrogenase and GMP reductase share many regions of sequence similarity. One of these regions is centered on a cysteine residue thought [3] to be involved in binding IMP. We have used this region as a signature pattern.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [LIVM]-[RK]-[LIVM]-G-[LIVM]-G-x-G-S-[LIVM]-C-x-T [C is the putative IMP-binding residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update May 1991 / First entry. References [1] Collart F.R., Huberman E. J. Biol. Chem. 263:15769-15772(1988).</p> <p>[2] Natsumeda Y., Ohno S., Kawasaki H., Konno Y., Weber G., Suzuki K. J. Biol. Chem. 265:5292-5295(1990).</p> <p>[3] Andrews S.C., Guest J.R. Biochem. J. 255:35-43(1988).</p>
Inos-1-P_synth		Myo-inositol-1-phosphate synthase	<p>Accession number: PF01658 Definition: Myo-inositol-1-phosphate synthase Author: Bashton M, Bateman A Alignment method of seed: Clustalw</p>

0066980 " 101300

Pfam	Prosite	Full Name	Description
			<p>Source of seed members: Pfam-B_959 (release 4.1)</p> <p>Gathering cutoffs: 25 25</p> <p>Trusted cutoffs: 86.80 86.80</p> <p>Noise cutoffs: -219.00 -219.00</p> <p>HMM build command line: hmmbuild -F HMM SEED</p> <p>HMM build command line: hmmlcalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 95066381</p> <p>Reference Title: Comparison of INO1 gene sequences and products in <i>Candida albicans</i> and <i>Saccharomyces cerevisiae</i>.</p> <p>Reference Author: Klig LS, Zobel PA, Devry CG, Losberger C;</p> <p>Reference Location: Yeast 1994;10:789-800.</p> <p>Database Reference INTERPRO; IPR002587;</p> <p>Comment: This is a family of myo-inositol-1-phosphate synthases.</p> <p>Comment: Inositol-1-phosphate catalyses the conversion of glucose-6-phosphate to inositol-1-phosphate, which is then dephosphorylated</p> <p>Comment: to inositol [1]. Inositol phosphates play an important role in</p> <p>Comment: signal transduction.</p> <p>Number of members: 27</p>
IPP_isomerase		Isopentenyl-diphosphate delta-isomerase	<p>Accession number: PF01772</p> <p>Definition: Isopentenyl-diphosphate delta-isomerase</p> <p>Author: Bashton M, Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_1099 (release 4.2)</p> <p>Gathering cutoffs: -88 -88</p> <p>Trusted cutoffs: -66.70 -66.70</p> <p>Noise cutoffs: -106.90 -106.90</p> <p>HMM build command line: hmmbuild -F HMM SEED</p> <p>HMM build command line: hmmlcalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 98409684</p> <p>Reference Title: Differential expression of two isopentenyl pyrophosphate isomerases and enhanced carotenoid accumulation in a unicellular chlorophyte</p> <p>Reference Author: Sun Z, Cunningham FX Jr, Gantt E;</p> <p>Reference Location: Proc Natl Acad Sci U S A 1998;95:11482-11488.</p> <p>Reference Number: [2]</p> <p>Reference Medline: 97373600</p> <p>Reference Title: Cloning and subcellular localization of hamster and rat isopentenyl diphosphate dimethylallyl diphosphate isomerase. A PTS1 motif targets the enzyme to peroxisomes.</p> <p>Reference Author: Paton VG, Shackelford JE, Krisans SK;</p> <p>Reference Location: J Biol Chem 1997;272:18945-18950.</p> <p>Database Reference INTERPRO; IPR002667;</p> <p>Comment: Isopentenyl-diphosphate delta-isomerase or IPP isomerase EC:5.3.3.2</p> <p>Comment: catalyses the interconversion of isopentenyl diphosphate and dimethylallyl diphosphate. Dimethylallyl phosphate is the initial substrate</p> <p>Comment: for the biosynthesis of carotenoids and other long chain isoprenoids [1].</p> <p>Number of members: 24</p>
K-box	PDOC00302	MADS-box domain signature and profile	<p>A number of transcription factors contain a conserved domain of 56 amino-acid residues, sometimes known as the MADS-box domain [E1]. They are listed below:</p> <p>- Serum response factor (SRF) [1], a mammalian transcription factor that</p>

0965980 "1013100

Pfam	Prosite	Full Name	Description
			[5] Sherman D.H., Malpartida F., Bibb M.J., Kieser H.M., Bibb M.J., Hopwood D.A. EMBO J. 8:2717-2725(1989).
KRAB		KRAB box	<p>Accession number: PF01352 Definition: KRAB box Author: Bateman A Alignment method of seed: Manual Source of seed members: Bateman A Gathering cutoffs: 0 0 Trusted cutoffs: 1.10 1.10 Noise cutoffs: -5.40 -5.40 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 91319563 Reference Title: Conserved KRAB protein domain identified upstream from the Reference Title: zinc finger region of Kox 8. Reference Author: Thiesen HJ, Bellefroid E, Revelant O, Martial JA; Reference Location: Nucleic Acids Res 1991;19:3996-3996. Reference Number: [2] Reference Medline: 97140325 Reference Title: A novel member of the RING finger family, KRIP-1, Reference Title: associates with the KRAB-A transcriptional repressor domain Reference Title: of zinc finger proteins. Reference Author: Kim SS, Chen YM, O'Leary E, Witzgall R, Vidal M, Bonventre Reference Author: JV; Reference Location: Proc Natl Acad Sci U S A 1996;93:15299-15304. Reference Number: [3] Reference Medline: 96365472 Reference Title: KAP-1, a novel corepressor for the highly conserved KRAB Reference Title: repression domain. Reference Author: Friedman JR, Fredericks WJ, Jensen DE, Speicher DW, Huang Reference Author: XP, Neilson EG, Rauscher FJ; Reference Location: Genes Dev 1996;10:2067-2078. Database Reference INTERPRO; IPR001909; Database reference: PFAMB; PB036541; Comment: The KRAB domain (or Kruppel-associated box) is present in Comment: about a third of zinc finger proteins containing C2H2 fingers. Comment: The KRAB domain is found to be involved in protein-protein Comment: interactions [2,3]. Comment: The KRAB domain is generally encoded by two exons. The Comment: regions coded by the two exons are known as KRAB-A and Comment: KRAB-B. Number of members: 105</p>
lectin_legB	PDOC00278	Legume lectins signatures	<p>Leguminous plants synthesize sugar-binding proteins which are called legume lectins [1,2]. These lectins are generally found in the seeds. The exact function of legume lectins is not known but they may be involved in the attachment of nitrogen-fixing bacteria to legumes and in the protection against pathogens. Legume lectins bind calcium and manganese (or other transition metals).</p> <p>Legume lectins are synthesized as precursor proteins of about</p>

09669980 101300

Pfam	Prosite	Full Name	Description
			<p>[1] Chapus C., Rovey M., Sarda L., Verger R. Biochimie 70:1223-1234(1988).</p> <p>[2] Persson B., Bengtsson-Olivecrona G., Enerback S., Olivecrona T., Joernvall H. Eur. J. Biochem. 179:39-45(1989).</p> <p>[3] Blow D. Nature 343:694-695(1990).</p> <p>[4] McLean J., Fielding C., Drayna D., Dieplinger H., Baer B., Kohr W., Henzel W., Lawn R. Proc. Natl. Acad. Sci. U.S.A. 83:2335-2339(1986).</p> <p>[5] Baker M.E. Biochem. J. 255:1057-1060(1988).</p>
Lipase_GDSL	PDOC00842	Lipolytic enzymes "G-D-S-L" family, serine active site	<p>Recently [1], a family of lipolytic enzymes has been characterized. This family currently consist of the following proteins:</p> <ul style="list-style-type: none"> - Aeromonas hydrophila lipase/phosphatidylcholine-sterol acyltransferase. - Xenorhabdus luminescens lipase 1. - Vibrio mimicus arylesterase. - Escherichia coli acyl-coA thioesterase I (gene tesA). - Vibrio parahaemolyticus thermolabile hemolysin/atypical phospholipase. - Rabbit phospholipase AdRab-B, an intestinal brush border protein with esterase and phospholipase A/lysophospholipase activity that could be involved in the uptake of dietary lipids. AdRab-B contains four repeats of about 320 amino acids. - Arabidopsis thaliana and Brassica napus anther-specific proline-rich protein APG. - A Pseudomonas putida hypothetical protein in trpE-trpG intergenic region. <p>A serine has been identified a part of the active site in the Aeromonas, Vibrio mimicus and Escherichia coli enzymes. It is located in a conserved sequence motif that can be used as a signature pattern for these proteins.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [LIVMFYAG](4)-G-D-S-[LIVM]-x(1,2)-[TAG]-G [S is the active site residue] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Note this pattern will pick up two of the four repeats in AdRab-B, the first one is not detected as its sequence has diverged in the region of the putative active site residue. The last one is also not detected because it is slightly divergent at the end of the pattern. Expert(s) to contact by email Upton C. upton@sol.uvic.ca</p> <p>Buckley J.T. tbuckley@sol.uvic.ca</p> <p>Last update November 1995 / First entry.</p>

09689980 "08668950

Pfam	Prosite	Full Name	Description
			<p>Reference Title: DNA-binding antibiotic and antitumour agent saframycin Mx1</p> <p>Reference Title: from <i>Myxococcus xanthus</i>.</p> <p>Reference Author: Pospiech A, Bietenhader J, Schupp T;</p> <p>Reference Location: Microbiology 1996;142:741-746.</p> <p>Database Reference: SCOP; 1vid; fa; [SCOP-USA][CATH-PDBSUM]</p> <p>Database Reference INTERPRO; IPR002935;</p> <p>Database Reference PDB; 1vid ; 13; 186;</p> <p>Database reference: PFAMB; PB040269;</p> <p>Comment: Members of this family are O-methyltransferases. The family</p> <p>Comment: includes catechol o-methyltransferase</p> <p>Swiss:P21964, caffeoyl-CoA</p> <p>Comment: O-methyltransferase Swiss:Q43095 and a</p> <p>family of bacterial</p> <p>Comment: O-methyltransferases that may be involved</p> <p>in antibiotic</p> <p>Comment: production [1].</p> <p>Number of members: 39</p>
MMR_HSR1		GTPase of unknown function	<p>Accession number: PF01926</p> <p>Definition: GTPase of unknown function</p> <p>Author: Enright A, Ouzounis C, Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Enright A</p> <p>Gathering cutoffs: -21 -21</p> <p>Trusted cutoffs: -20.70 -20.70</p> <p>Noise cutoffs: -31.60 -31.60</p> <p>HMM build command line: hmmbuild HMM SEED</p> <p>HMM build command line: hmmcalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 94235953</p> <p>Reference Title: Structure and evolution of a member of a new subfamily of</p> <p>Reference Title: GTP-binding proteins mapping to the</p> <p>human MHC class I</p> <p>Reference Title: region.</p> <p>Reference Author: Vernet C, Ribouchon MT, Chimini</p> <p>GPontarotti P;</p> <p>Reference Location: Mamm Genome 1994;5:100-105.</p> <p>Database Reference INTERPRO; IPR002917;</p> <p>Database reference: PFAMB; PB000471;</p> <p>Database reference: PFAMB; PB002171;</p> <p>Database reference: PFAMB; PB015790;</p> <p>Number of members: 67</p>
MoaC		MoaC family	<p>Accession number: PF01967</p> <p>Definition: MoaC family</p> <p>Author: Enright A, Ouzounis C, Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Enright A</p> <p>Gathering cutoffs: 25 25</p> <p>Trusted cutoffs: 73.00 73.00</p> <p>Noise cutoffs: -93.90 -93.90</p> <p>HMM build command line: hmmbuild -F HMM SEED</p> <p>HMM build command line: hmmcalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 99337076</p> <p>Reference Title: Characterization of a molybdenum cofactor biosynthetic gene</p> <p>Reference Title: cluster in <i>Rhodobacter capsulatus</i> which is specific for the</p> <p>Reference Title: biogenesis of dimethylsulfoxide reductase.</p> <p>Reference Author: Solomon PS, Shaw AL, Lane I, Hanson</p> <p>GR, Palmer T, McEwan</p> <p>Reference Author: AG;</p> <p>Reference Location: Microbiology 1999;145:1421-1429.</p> <p>Database Reference INTERPRO; IPR002820;</p> <p>Comment: Members of this family are involved in molybdenum</p> <p>Comment: cofactor biosynthesis. However their</p> <p>molecular</p> <p>Comment: function is not known.</p>

0068980 "101300

Pfam	Prosite	Full Name	Description
			<p>- The melibiose carrier (gene melB) from a variety of enterobacteria. This protein is responsible for melibiose transport and is capable of using hydrogen, sodium, and lithium cations as coupling cations for cotransport.</p> <p>- The lactose permease from <i>Lactobacillus</i> (gene lacS or lacY). This protein is responsible for the transport of beta-galactosides into the cell, with the concomitant export of a proton. It consists of two domains; a N-terminal SGF domain and a C-terminal domain that resembles that of enzyme IIA of the PEP:sugar phosphotransferase system.</p> <p>- The raffinose permease from <i>Pediococcus pentosaceus</i>. It also consists of a N-terminal SGF domain and a C-terminal IIA domain.</p> <p>- The glucuronide carrier (gene gusB or uidP) from <i>Escherichia coli</i>.</p> <p>- The xylose transporter (gene xylP) from <i>Lactobacillus pentosus</i>.</p> <p>- <i>Escherichia coli</i> hypothetical protein yagG.</p> <p>- <i>Escherichia coli</i> hypothetical protein yicJ.</p> <p>- <i>Escherichia coli</i> hypothetical protein yihO.</p> <p>- <i>Escherichia coli</i> hypothetical protein yihP.</p> <p>- <i>Bacillus subtilis</i> hypothetical protein yjmB.</p> <p>- <i>Bacillus subtilis</i> hypothetical protein ynaJ.</p> <p>Like sugar transport proteins, these integral membrane proteins are predicted to comprise twelve membrane spanning domains. As a signature pattern we selected a highly conserved region which is located in a cytoplasmic loop between the second and third transmembrane regions. This region starts with a conserved aspartate which has been shown [2], in melB, to be important for the activity of the protein.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [DG]-x(3)-G-x(3)-[DN]-x(6,8)-[GA]-[KRHQ]-[FSA]-[KR]-[PT]-[FYW]-[LIVMWQ]-[LIV]-x-[GAFV]-[GSTA]</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update July 1999 / Pattern and text revised.</p> <p>References [1] Reizer J., Reizer A., Saier M.H. Jr. <i>Biochim. Biophys. Acta</i> 1197:133-136(1994).</p> <p>[2] Pourcher T., Deckert M., Bassilana M., Leblanc G. <i>Biochem. Biophys. Res. Commun.</i> 178:1176-1181(1991).</p>
Na_K_ATPase_C		Na+/K+ ATPase C-terminus	<p>This domain is specific to the sodium and potassium ATPases (Na_K-ATPase).</p> <p>The sodium pump (Na+,K+ ATPase), located in the plasma membrane of all animal cells [1], is a heterotrimer of a catalytic subunit (alpha chain), a glycoprotein subunit of about 34 Kd (beta chain) and a small hydrophobic protein of about 6 Kd. The beta subunit seems [2] to regulate, through the assembly of alpha/beta heterodimers, the number of sodium pumps transported to the plasma membrane.</p> <p>This family is typically found in association with E1-E2 ATPase. Uses of these polypeptide includes regulating that ion content in a desired cell or organism and can convey salt or ion tolerance.</p>
Na_K_ATPase_N		Na+/K+ ATPase C-	Accession number: PF00690

09668960 "08668960"

Pfam	Prosite	Full Name	Description
			<p>catalyses the transfer of two electrons from NADH to ubiquinone in a reaction that is associated with proton translocation across the membrane.</p> <p>Number of members: 546</p>
oxidored_q2		NADH-ubiquinone/plastoquinone oxidoreductase chain 4L	<p>Accession number: PF00420 Definition: NADH-ubiquinone/plastoquinone oxidoreductase chain 4L Author: Finn RD Alignment method of seed: Clustalw Source of seed members: Pfam-B_193 (release 1.0) Gathering cutoffs: 25 15 Trusted cutoffs: 29.70 29.70 Noise cutoffs: 20.40 20.40 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmbuild --seed 0 HMM Database Reference: INTERPRO; IPR001133; Database reference: PFAMB; PB006066; Number of members: 219</p>
PAN	PDOC00376	Apple domain	<p>Plasma kallikrein (EC 3.4.21.34) and coagulation factor XI (EC 3.4.21.27) are two related plasma serine proteases activated by factor XIIA and which share the same domain topology: an N-terminal region that contains four tandem repeats of about 90 amino acids and a C-terminal catalytic domain.</p> <p>The 90 amino-acid repeated domain contains 6 conserved cysteines. It has been shown [1,2] that three disulfide bonds link the first and sixth, second and fifth, and third and fourth cysteines. The domain can be drawn in the shape of an apple (see below) and has been accordingly called the 'apple domain'.</p> <pre> x x x x x x x C---C x x x x x x C x x x x x x x x C x x x x x x x x x x x x x x x x x x x x x x C---C x x..... </pre> <p>Schematic representation of an apple domain.</p> <p>Apart from the cysteines, there are a number of other conserved positions in the apple domain. We have developed a pattern, that spans the complete domain, and which includes these conserved positions.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern C-x(3)-[LIVMFY]-x(5)-[LIVMFY]-x(3)-[DENQ]-[LIVMFY]-x(10)-C-x(3)-C-T-x(4)-C-x-[LIVMFY]-F-x-[FY]-x(13,14)-C-x-[LIVMFY]-[RK]-x-[ST]-x(14,15)-S-G-x-[ST]-[LIVMFY]-x(2)-C</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update June 1992 / Pattern and text revised.</p> <p>References</p>

00689980 101300

Pfam	Prosite	Full Name	Description
			<p>Reference Title: ATP sulphurylase activity of the nodP and nodQ gene</p> <p>Reference Title: products of Rhizobium meliloti.</p> <p>Reference Author: Schwedock J, Long SR;</p> <p>Reference Location: Nature 1990;348:644-647.</p> <p>Database Reference: SCOP; 1sur; fa; [SCOP-USA][CATH-PDBSUM]</p> <p>Database Reference INTERPRO; IPR002500;</p> <p>Database Reference PDB; 1sur ; 48; 215;</p> <p>Comment: This domain is found in phosphoadenosine phosphosulfate (PAPS) reductase</p> <p>Comment: enzymes or PAPS sulfotransferase. PAPS reductase is part of the adenine</p> <p>Comment: nucleotide alpha hydrolases superfamily also including N type ATP PPases</p> <p>Comment: and ATP sulphurylases [1]. The enzyme uses thioredoxin as an electron</p> <p>Comment: donor for the reduction of PAPS to phospho-adenosine-phosphate (PAP) [1,2].</p> <p>Comment: It is also found in NodP nodulation protein P from Rhizobium which has ATP</p> <p>Comment: sulphurylase activity (sulfate adenylate transferase) [3].</p> <p>Number of members: 48</p>
PARP		Poly(ADP-ribose) polymerase catalytic region	<p>Accession number: PF00644</p> <p>Definition: Poly(ADP-ribose) polymerase catalytic region.</p> <p>Author: Bateman A</p> <p>Alignment method of seed: HMM_built_from_alignment</p> <p>Source of seed members: Bateman A</p> <p>Gathering cutoffs: -59.4 -59.4</p> <p>Trusted cutoffs: -44.60 -44.60</p> <p>Noise cutoffs: -180.60 -180.60</p> <p>HMM build command line: hmmbuild HMM SEED</p> <p>HMM build command line: hmmcalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 96353841</p> <p>Reference Title: Structure of the catalytic fragment of poly(AD-ribose)</p> <p>Reference Title: polymerase from chicken.</p> <p>Reference Author: Ruf A, Mennissier de Murcia J, de Murcia G, Schulz GE;</p> <p>Reference Location: Proc Natl Acad Sci U S A 1996;93:7481-7485.</p> <p>Reference Number: [2]</p> <p>Reference Medline: 93293867</p> <p>Reference Title: The carboxyl-terminal domain of human poly(ADP-ribose)</p> <p>Reference Title: polymerase. Overproduction in Escherichia coli, large scale</p> <p>Reference Title: purification, and characterization.</p> <p>Reference Author: Simonin F, Hofferer L, Panzeter PL, Muller S, de Murcia G,</p> <p>Reference Author: Althaus FR;</p> <p>Reference Location: J Biol Chem 1993;268:13454-13461.</p> <p>Database Reference: SCOP; 1paw; fa; [SCOP-USA][CATH-PDBSUM]</p> <p>Database Reference INTERPRO; IPR001290;</p> <p>Database Reference PDB; 1a26 ; 662; 997;</p> <p>Database Reference PDB; 1pax ; 662; 997;</p> <p>Database Reference PDB; 2pax ; 662; 997;</p> <p>Database Reference PDB; 3pax ; 662; 997;</p> <p>Database Reference PDB; 4pax ; 662; 997;</p> <p>Database Reference PDB; 2paw ; 662; 1009;</p> <p>Database reference: PFAMB; PB041409;</p> <p>Comment: Poly(ADP-ribose) polymerase catalyses the covalent</p> <p>Comment: attachment of ADP-ribose units from NAD+ to itself and</p> <p>Comment: to a limited number of other DNA binding proteins, which</p> <p>Comment: decreases their affinity for DNA.</p> <p>Comment: Poly(ADP-ribose) polymerase is a regulatory component</p>

0066980 " 0866960

Pfam	Prosite	Full Name	Description
			<p>Comment: induced by DNA damage.</p> <p>Comment: The carboxyl-terminal region is the most highly conserved</p> <p>Comment: region of the protein. Experiments have shown that a</p> <p>Comment: carboxyl 40 kDa fragment is still catalytically active [2].</p> <p>Number of members: 19</p>
PC_rep		Proteasome/cyclosome repeat	<p>Accession number: PF01851</p> <p>Definition: Proteasome/cyclosome repeat</p> <p>Author: Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: [1]</p> <p>Gathering cutoffs: 25 0</p> <p>Trusted cutoffs: 30.60 3.00</p> <p>Noise cutoffs: 15.80 15.80</p> <p>HMM build command line: hmmbuild HMM SEED</p> <p>HMM build command line: hmmscalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 97348748</p> <p>Reference Title: A repetitive sequence in subunits of the 26S proteasome and</p> <p>Reference Title: 20S cyclosome (anaphase-promoting complex).</p> <p>Reference Author: Lupas A, Baumeister W, Hofmann K;</p> <p>Reference Location: Trends Biochem Sci 1997;22:195-196.</p> <p>Database Reference: INTERPRO; IPR002015;</p> <p>Database reference: PFAMB; PB009978;</p> <p>Database reference: PFAMB; PB040656;</p> <p>Number of members: 112</p>
PE		PE family	<p>Accession number: PF00934</p> <p>Definition: PE family</p> <p>Author: Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_253 (release 3.0)</p> <p>Gathering cutoffs: -20 -20</p> <p>Trusted cutoffs: -10.80 -10.80</p> <p>Noise cutoffs: -20.60 -20.60</p> <p>HMM build command line: hmmbuild HMM SEED</p> <p>HMM build command line: hmmscalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 98295987</p> <p>Reference Title: Deciphering the biology of Mycobacterium tuberculosis from</p> <p>Reference Title: the complete genome sequence.</p> <p>Reference Author: Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C,</p> <p>Reference Author: Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd,</p> <p>Reference Author: Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T,</p> <p>Reference Author: Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin</p> <p>Reference Author: N, Holroyd S, Hornsby T, Jagels K, Barrell BG, et al;</p> <p>Reference Location: Nature 1998;393:537-544..</p> <p>Database Reference: INTERPRO; IPR000084;</p> <p>Comment: This family named after a PE motif near to the amino</p> <p>Comment: terminus of the domain. The PE family of proteins</p> <p>Comment: all contain an amino-terminal region of about 110</p> <p>Comment: amino acids. The carboxyl terminus of this family</p> <p>Comment: are variable and fall into several classes. The</p> <p>Comment: largest class of PE proteins is the highly repetitive</p> <p>Comment: PGRS class which have a high glycine content.</p> <p>Comment: The function of these proteins is uncertain</p>

00669980-101300

Pfam	Prosite	Full Name	Description
			<p>These proteases participates in the breakdown of the egg envelope, which is derived from the egg extracellular matrix, at the time of hatching.</p> <p>Family M12B</p> <ul style="list-style-type: none"> - Snake venom metalloproteinases [6]. This subfamily mostly groups proteases that act in hemorrhage. Examples are: adamalysin II (EC 3.4.24.46), atrolysin C/D (EC 3.4.24.42), atrolysin E (EC 3.4.24.44), fibrolase (EC 3.4.24.72), trimerelysin I (EC 3.4.25.52) and II (EC 3.4.25.53). - Mouse cell surface antigen MS2. <p>Family M13</p> <ul style="list-style-type: none"> - Mammalian neprilysin (EC 3.4.24.11) (neutral endopeptidase) (NEP). - Endothelin-converting enzyme 1 (EC 3.4.24.71) (ECE-1), which process the precursor of endothelin to release the active peptide. - Kell blood group glycoprotein, a major antigenic protein of erythrocytes. <p>The Kell protein is very probably a zinc endopeptidase.</p> <ul style="list-style-type: none"> - Peptidase O from <i>Lactococcus lactis</i> (gene pepO). <p>Family M27</p> <ul style="list-style-type: none"> - Clostridial neurotoxins, including tetanus toxin (TeTx) and the various botulinum toxins (BoNT). These toxins are zinc proteases that block neurotransmitter release by proteolytic cleavage of synaptic proteins such as synaptobrevins, syntaxin and SNAP-25 [7,8]. <p>Family M30</p> <ul style="list-style-type: none"> - <i>Staphylococcus hyicus</i> neutral metalloprotease. <p>Family M32</p> <ul style="list-style-type: none"> - Thermostable carboxypeptidase 1 (EC 3.4.17.19) (carboxypeptidase Taq), an enzyme from <i>Thermus aquaticus</i> which is most active at high temperature. <p>Family M34</p> <ul style="list-style-type: none"> - Lethal factor (LF) from <i>Bacillus anthracis</i>, one of the three proteins composing the anthrax toxin. <p>Family M35</p> <ul style="list-style-type: none"> - Deuterolysin (EC 3.4.24.39) from <i>Penicillium citrinum</i> and related proteases from various species of <i>Aspergillus</i>. <p>Family M36</p> <ul style="list-style-type: none"> - Extracellular elastinolytic metalloproteinases from <i>Aspergillus</i>. <p>From the tertiary structure of thermolysin, the position of the residues acting as zinc ligands and those involved in the catalytic activity are known. Two of the zinc ligands are histidines which are very close together in the sequence; C-terminal to the first histidine is a glutamic acid residue which acts as a nucleophile and promotes the attack of a water molecule on the carbonyl carbon of the substrate. A signature pattern which includes the two histidine and the glutamic acid residues is sufficient to detect this superfamily of proteins.</p>

09589980 101300

Pfam	Prosite	Full Name	Description
			<ul style="list-style-type: none"> - Kexin (EC 3.4.21.61) from yeast (gene KEX-2). - Oryzin (EC 3.4.21.63) (alkaline proteinase) from <i>Aspergillus</i> (gene alp). - Proteinase K (EC 3.4.21.64) from <i>Tritirachium album</i> (gene proK). - Proteinase R from <i>Tritirachium album</i> (gene proR). - Proteinase T from <i>Tritirachium album</i> (gene proT). - Subtilisin-like protease III from yeast (gene YSP3). - Thermomycolin (EC 3.4.21.65) from <i>Malbranchea sulfurea</i>. <p>- Furin (EC 3.4.21.85), neuroendocrine convertases 1 to 3 (NEC-1 to -3) and PACE4 protease from mammals, other vertebrates, and invertebrates. These proteases are involved in the processing of hormone precursors at sites comprised of pairs of basic amino acid residues [3].</p> <ul style="list-style-type: none"> - Tripeptidyl-peptidase II (EC 3.4.14.10) (tripeptidyl aminopeptidase) from Human. - Prestalk-specific proteins tagB and tagC from slime mold [4]. Both proteins consist of two domains: a N-terminal subtilase catalytic domain and a C-terminal ABC transporter domain (see <PDOC00185>). <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [STAIV]-x-[LIVMF]-[LIVM]-D-[DSTA]-G-[LIVMFC]-x(2,3)-[DNH] [D is the active site residue] Sequences known to belong to this class detected by the pattern the majority of subtilases with a few exceptions. Other sequence(s) detected in SWISS-PROT 44.</p> <p>Consensus pattern H-G-[STM]-x-[VIC]-[STAGC]-[GS]-x-[LIVMA]-[STAGCLV]-[SAGM] [H is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for aspA and ssa1 which both seem to lack the histidine active site. Other sequence(s) detected in SWISS-PROT adenylate cyclase type VIII.</p> <p>Consensus pattern G-T-S-x-[SA]-x-P-x(2)-[STAVC]-[AG] [S is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for nisP, tagC and <i>S.marcescens</i> extracellular serine protease. Other sequence(s) detected in SWISS-PROT 6.</p> <p>Note if a protein includes at least two of the three active site signatures, the probability of it being a serine protease from the subtilase family is 100%</p> <p>Note these proteins belong to family S8 in the classification of peptidases [5,E1]. Expert(s) to contact by email Brannigan J. jab5@vaxa.york.ac.uk</p> <p>Siezen R.J. siezen@nizo.nl</p> <p>Last update November 1997 / Patterns and text revised. References [1] Siezen R.J., de Vos W.M., Leunissen J.A.M., Dijkstra B.W. <i>Protein Eng.</i> 4:719-737(1991). [2] Siezen R.J. (In) <i>Proceeding subtilisin symposium, Hamburg, (1992).</i> [3]</p>

09689980 101300

Pfam	Prosite	Full Name	Description
			<p>Barr P.J. Cell 66:1-3(1991).</p> <p>[4] Shaulsky G., Kuspa A., Loomis W.F.; Genes Dev. 9:1111-1122(1995).</p> <p>[5] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).</p> <p>[E1] http://www.expasy.ch/cgi-bin/lists?peptidas.txt</p>
Peptidase_S9	PDOC00587	Prolyl oligopeptidase family serine active site	<p>The prolyl oligopeptidase family [1,2,3] consist of a number of evolutionary related peptidases whose catalytic activity seems to be provided by a charge relay system similar to that of the trypsin family of serine proteases, but which evolved by independent convergent evolution. The known members of this family are listed below.</p> <ul style="list-style-type: none"> - Prolyl endopeptidase (EC 3.4.21.26) (PE) (also called post-proline cleaving enzyme). PE is an enzyme that cleaves peptide bonds on the C-terminal side of prolyl residues. The sequence of PE has been obtained from a mammalian species (pig) and from bacteria (Flavobacterium meningosepticum and Aeromonas hydrophila); there is a high degree of sequence conservation between these sequences. - Escherichia coli protease II (EC 3.4.21.83) (oligopeptidase B) (gene prtB) which cleaves peptide bonds on the C-terminal side of lysyl and arginyl residues. - Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV). DPP IV is an enzyme that removes N-terminal dipeptides sequentially from polypeptides having unsubstituted N-termini provided that the penultimate residue is proline. - Yeast vacuolar dipeptidyl aminopeptidase A (DPAP A) (gene: STE13) which is responsible for the proteolytic maturation of the alpha-factor precursor. - Yeast vacuolar dipeptidyl aminopeptidase B (DPAP B) (gene: DAP2). - Acylamino-acid-releasing enzyme (EC 3.4.19.1) (acyl-peptide hydrolase). This enzyme catalyzes the hydrolysis of the amino-terminal peptide bond of an N-acetylated protein to generate a N-acetylated amino acid and a protein with a free amino-terminus. <p>A conserved serine residue has experimentally been shown (in E.coli protease II as well as in pig and bacterial PE) to be necessary for the catalytic mechanism. This serine, which is part of the catalytic triad (Ser, His, Asp), is generally located about 150 residues away from the C-terminal extremity of these enzymes (which are all proteins that contains about 700 to 800 amino acids).</p>

00669900 101300

Pfam	Prosite	Full Name	Description																
			<p>they consist of a single polypeptide chain (called pilin or fimbrial protein) arranged in a helical configuration of five subunits per turn in the assembled pilus. Gram-negative bacteria produce pilin which are characterized by the presence of a very short leader peptide of 6 to 7 residues, followed by a methylated N-terminal phenylalanine residue and by a highly conserved sequence of about 24 hydrophobic residues. This class of pilin is often referred to as NMePhe or type-4 pili [1,2].</p> <p>Recently a number of bacterial proteins have been sequenced which share the following structural characteristics with type-4 pili [3]:</p> <p>a) The N-terminal residue, which is methylated, is hydrophobic (generally a phenylalanine or a methionine);</p> <p>b) The leader peptide is hydrophilic, consists of 5 to 10 residues (with two exceptions, see below) and ends with a glycine;</p> <p>c) The fifth residue of the mature sequence is a glutamate which seems to be required for the methylation step;</p> <p>d) The first twenty residues of the mature sequence are highly hydrophobic.</p> <p>These proteins are listed below:</p> <p>- Four proteins in an operon involved in a general secretion pathway (GSP) for the export of proteins (also called the type II pathway) [4]. These proteins have been assigned a different gene name in each of the species where they have been sequenced:</p> <table><tr><th>Species</th><th>Gene names</th></tr><tr><td>Aeromonas hydrophila</td><td>exeG exeH exeI exeJ</td></tr><tr><td>Erwinia chrysanthemi</td><td>outG outH outI outJ</td></tr><tr><td>Escherichia coli</td><td>hofG hofH yheH yheI</td></tr><tr><td>Klebsiella pneumoniae</td><td>pulG pulH pulI pulJ</td></tr><tr><td>Pseudomonase aeruginosa</td><td>xcpT xcpU xcpV xcpW</td></tr><tr><td>Vibrio cholerae</td><td>epsG epsH epsI epsJ</td></tr><tr><td>Xanthomonas campestris</td><td>xpsG xpsH xpsI xpsJ</td></tr></table> <p>- Vibrio cholerae toxin co-regulated pilin (gene tcpA). This pilin has a much longer putative leader peptide (25 residues).</p> <p>- Bacillus subtilis comG competence operon proteins 3, 4, and 5 which are involved for the uptake of DNA by competent Bacillus subtilis cells.</p> <p>- ppdA, ppdB and ppdC, three Escherichia coli hypothetical proteins found in the thyA-recC intergenic region.</p> <p>- ppdA, a hypothetical protein near the groeLS operon of Clostridium perfringens. The putative leader peptide is 23 residues long.</p> <p>We developed a signature pattern based on the N-terminal conserved region of all these proteins.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [KRHEQSTAG]-G-[FYLIVM]-[ST]-[LT]-[LIVP]-E-[LIVFWSTAG](14) [The residue after the G is methylated]</p>	Species	Gene names	Aeromonas hydrophila	exeG exeH exeI exeJ	Erwinia chrysanthemi	outG outH outI outJ	Escherichia coli	hofG hofH yheH yheI	Klebsiella pneumoniae	pulG pulH pulI pulJ	Pseudomonase aeruginosa	xcpT xcpU xcpV xcpW	Vibrio cholerae	epsG epsH epsI epsJ	Xanthomonas campestris	xpsG xpsH xpsI xpsJ
Species	Gene names																		
Aeromonas hydrophila	exeG exeH exeI exeJ																		
Erwinia chrysanthemi	outG outH outI outJ																		
Escherichia coli	hofG hofH yheH yheI																		
Klebsiella pneumoniae	pulG pulH pulI pulJ																		
Pseudomonase aeruginosa	xcpT xcpU xcpV xcpW																		
Vibrio cholerae	epsG epsH epsI epsJ																		
Xanthomonas campestris	xpsG xpsH xpsI xpsJ																		

00669980 "08669980

Pfam	Prosite	Full Name	Description
			<p>HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild --seed 0 HMM Reference Number: [1] Reference Medline: 94233771 Reference Title: Changes in the amino acid sequence of the coat protein Reference Title: readthrough domain of potato leafroll Reference Title: luteovirus affect the formation of an epitope and aphid transmission. Reference Author: Jolly CA, Mayo MA; Reference Location: Virology 1994;201:182-185. Database Reference INTERPRO; IPR002929; Comment: This family consists mainly of the potato leaf roll virus Comment: readthrough protein. This is generated via a readthrough Comment: of open reading frame 3 a coat protein allowing transcription Comment: of open reading frame 5 to give an extended coat protein Comment: with a large c-terminal addition or read through domain [1]. Comment: The readthrough protein is thought to play a role in the Comment: circulative aphid transmission of potato leaf roll virus [1]. Comment: Also in the family is open reading frame 6 from beet western Comment: yellows virus and potato leaf roll virus both luteovirus and Comment: an unknown protein from cucurbit aphid-borne yellows virus a Comment: closterovirus. Number of members: 28</p>
PMSR		Peptide methionine sulfoxide reductase	<p>Accession number: PF01625 Definition: Peptide methionine sulfoxide reductase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_1111 (release 4.1) Gathering cutoffs: -62 -62 Trusted cutoffs: -28.00 -28.00 Noise cutoffs: -96.70 -96.70 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmbuild --seed 0 HMM Reference Number: [1] Reference Medline: 96353931 Reference Title: Peptide methionine sulfoxide reductase Reference Title: contributes to the maintenance of adhesins in three major pathogens. Reference Author: Wizemann TM, Moskovitz J, Pearce BJ, Cundell D, Arvidson Reference Author: CG, So M, Weissbach H, Brot N, Masure HR; Reference Location: Proc Natl Acad Sci USA 1996;93:7985-7990. Reference Number: [2] Reference Medline: 96312545 Reference Title: Cloning the expression of a mammalian gene involved in the Reference Title: reduction of methionine sulfoxide residues in proteins. Reference Author: Moskovitz J, Weissbach H, Brot N; Reference Location: Proc Natl Acad Sci U S A 1996;93:2095-2099. Database Reference INTERPRO; IPR002569; Comment: This enzyme repairs damaged proteins. Methionine sulfoxide in proteins Comment: is reduced to methionine. Number of members: 28</p>
Pollen allerg 2		Ribonuclease (pollen)	<p>Accession number: PF01620</p>

09589980 101300

Pfam	Prosite	Full Name	Description
		allergen)	<p>Definition: Ribonuclease (pollen allergen)</p> <p>Author: Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_1050 (release 4.1)</p> <p>Gathering cutoffs: -3 -3</p> <p>Trusted cutoffs: 23.10 23.10</p> <p>Noise cutoffs: -29.40 -29.40</p> <p>HMM build command line: hmmbuild -F HMM SEED</p> <p>HMM build command line: hmmscalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 95246885</p> <p>Reference Title: Major allergen Phl p Vb in timothy grass is a novel pollen</p> <p>Reference Author: Bufe A, Schramm G, Keown MB, Schlaak M, Becker WM;</p> <p>Reference Location: Febs lett 1995;363:6-12.</p> <p>Database Reference: INTERPRO; IPR002914;</p> <p>Database reference: PFAMB; PB037130;</p> <p>Comment: This family contains grass pollen proteins of group V.</p> <p>Comment: Swiss:Q40963 has been shown to possess ribonuclease</p> <p>Comment: activity [1].</p> <p>Number of members: 27</p>
POR_N		Pyruvate flavodoxin/ferredoxin oxidoreductase (N terminus)	<p>Accession number: PF01855</p> <p>Definition: Pyruvate flavodoxin/ferredoxin oxidoreductase (N terminus)</p> <p>Author: Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_323 (release 4.2)</p> <p>Gathering cutoffs: -116 -116</p> <p>Trusted cutoffs: -113.60 -113.60</p> <p>Noise cutoffs: -119.50 -119.50</p> <p>HMM build command line: hmmbuild -F HMM SEED</p> <p>HMM build command line: hmmscalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 96125254</p> <p>Reference Title: Molecular and phylogenetic characterization of pyruvate and</p> <p>Reference Title: 2-ketoisovalerate ferredoxin oxidoreductases from</p> <p>Reference Title: Pyrococcus furiosus and pyruvate ferredoxin oxidoreductase</p> <p>Reference Title: from Thermotoga maritima.</p> <p>Reference Author: Kletzin A, Adams MW;</p> <p>Reference Location: J Bacteriol 1996;178:248-257.</p> <p>Reference Number: [2]</p> <p>Reference Medline: 94022264</p> <p>Reference Title: Growth of the cyanobacterium Anabaena on molecular</p> <p>Reference Title: nitrogen: NifJ is required when iron is limited.</p> <p>Reference Author: Bauer CC, Scappino L, Haselkorn R;</p> <p>Reference Location: Proc Natl Acad Sci U S A 1993;90:8812-8816.</p> <p>Reference Number: [3]</p> <p>Reference Medline: 99140300</p> <p>Reference Title: Crystal structures of the key anaerobic enzyme</p> <p>Reference Title: pyruvate:ferredoxin oxidoreductase, free and in complex</p> <p>Reference Title: with pyruvate.</p> <p>Reference Author: Chabriere E, Charon MH, Volbeda A, Pieulle L, Hatchikian</p> <p>Reference Author: EC, Fontecilla-Camps JC;</p> <p>Reference Location: Nat Struct Biol 1999;6:182-190.</p> <p>Database Reference: SCOP; 2pda; fa; [SCOP-USA][CATH-PDBSUM]</p> <p>Database Reference: SCOP; 2pda; fa; [SCOP-USA][CATH-PDBSUM]</p> <p>Database Reference: INTERPRO; IPR002880;</p> <p>Database Reference: PDB; 1b0p A; 43; 328;</p>

00689980 "013000

Pfam	Prosite	Full Name	Description
			<p>Reference Medline: 10322432 Reference Title: The PWI motif: a new protein domain in splicing factors. Reference Author: Blencowe BJ, Ouzounis CA; Reference Location: Trends Biochem Sci 1999;24:179-180. Database Reference: INTERPRO; IPR002483; Number of members: 11</p>
R3H		R3H domain	<p>Accession number: PF01424 Definition: R3H domain Author: Bateman A Alignment method of seed: Manual Source of seed members: Medline:99003905 Gathering cutoffs: 25 25 Trusted cutoffs: 59.30 59.30 Noise cutoffs: 5.10 5.10 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 99003905 Reference Title: The R3H motif: a domain that binds single-stranded nucleic acids. Reference Author: Grishin NV; Reference Location: Trends Biochem Sci 1998;23:329-330. Database Reference: INTERPRO; IPR001374; Database reference: PFAMB; PB041444; Comment: The name of the R3H domain comes from the characteristic spacing Comment: of the most conserved arginine and histidine residues. The Comment: function of the domain is predicted to be binding ssDNA. Number of members: 28</p>
RepB_protein		Initiator RepB protein	<p>Accession number: PF01051 Definition: Initiator RepB protein Author: Finn RD, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_313 (release 3.0) Gathering cutoffs: 14 14 Trusted cutoffs: 19.00 16.20 Noise cutoffs: 11.80 12.90 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98284148 Reference Title: Replication and control of circular bacterial plasmids. Reference Author: del Solar G, Giraldo R, Ruiz-Echevarria MJ, Espinosa M, Reference Author: Diaz-Orejas R; Reference Location: Microbiol Mol Biol Rev 1998;62:434-464. Reference Number: [2] Reference Medline: 97324207 Reference Title: Initiation of replication of plasmid pMV158: mechanisms of Reference Title: DNA strand-transfer reactions mediated by the initiator Reference Title: RepB protein. Reference Author: Moscoso M, Eritja R, Espinosa M; Reference Location: J Mol Biol 1997;268:840-856. Database Reference: INTERPRO; IPR000525; Database Reference: PDB; 1rep C; 198; 240; Database reference: PFAMB; PB000509; Comment: This protein is an initiator of plasmid replication. Comment: RepB possesses nicking-closing (topoisomerase I) like activity. Comment: It is also able to perform a strand transfer reaction on ssDNA Comment: that contains its target. Number of members: 51</p>

09689980 "101300

Pfam	Prosite	Full Name	Description
			<p>- Halobacterium marismortui HL30 [2].</p> <p>These proteins have 87 to 128 amino-acid residues. As a signature pattern, we selected a conserved region located in the central section.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern V-[KR]-[LIVM]-x(3)-[LIVM]-N-x-[AKH]-x-W-x-[KR]-G</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update July 1999 / Pattern and text revised.</p> <p>References [1] Tanaka T., Kuwano Y., Kuzumaki T., Ishikawa K., Ogata K. Eur. J. Biochem. 162:45-48(1987).</p> <p>[2] Bergmann U., Arndt E. Biochim. Biophys. Acta 1050:56-60(1990).</p>
Ribosomal_L35Ae	PDOC00849	Ribosomal protein L35Ae signature	<p>A number of eukaryotic and archaebacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:</p> <ul style="list-style-type: none"> - Vertebrate L35A. - Caenorhabditis elegans L35A (F10E7.7). - Yeast L37A/L37B (Rp47). - Pyrococcus woesei L35A homolog [1]. <p>These proteins have 87 to 110 amino-acid residues. As a signature pattern, we selected a highly conserved stretch of 22 residues in the C-terminal part of these proteins.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern G-K-[LIVM]-x-R-x-H-G-x(2)-G-x-V-x-A-x-F-x(3)-[LI]-P</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update November 1997 / Pattern and text revised.</p> <p>References [1] Ouzounis C., Kyripides N., Sander C. Nucleic Acids Res. 23:565-570(1995).</p>
Ribosomal_L35p	PDOC00721	Ribosomal protein L35 signature	<p>Ribosomal protein L35 is one of the proteins from the large subunit of the ribosome. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups:</p> <ul style="list-style-type: none"> - Eubacterial L35. - Plant chloroplast L35 (nuclear-encoded). - Red algal chloroplast L35. - Cyanelle L35. <p>L35 is a basic protein of 60 to 70 amino-acid residues. As a signature pattern we selected a conserved region in the N-terminal section.</p>

09689980-101300

Pfam	Prosite	Full Name	Description
			<p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [LIVM]-K-[TV]-x(2)-[GSA]-[SAILV]-x-K-R-[LIVMFY]-[KRLS]</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update December 1999 / Pattern and text revised.</p> <p>References [1] Otaka E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).</p>
Ribosomal_L36e	PDOC00916	Ribosomal protein L36e signature	<p>A number of eukaryotic ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:</p> <ul style="list-style-type: none"> - Mammalian L36 [1]. - Drosophila L36 (M(1)1B). - Caenorhabditis elegans L36 (F37C12.4). - Candida albicans L39. - Yeast YL39. <p>These proteins have 99 to 104 amino acids. As a signature pattern, we selected a conserved region in the central part of these proteins.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern P-Y-E-[KR]-R-x-[LIVM]-[DE]-[LIVM](2)-[KR]</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update November 1997 / First entry.</p> <p>References [1] Chan Y.-L., Paz V., Olivera J., Wool I.G. Biochem. Biophys. Res. Commun. 192:849-853(1993).</p>
Ribosomal_L37ae		Ribosomal L37ae protein family	<p>Accession number: PF01780</p> <p>Definition: Ribosomal L37ae protein family</p> <p>Author: Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: PSI-BLAST P54051</p> <p>Gathering cutoffs: 25 25</p> <p>Trusted cutoffs: 145.10 145.10</p> <p>Noise cutoffs: -46.90 -46.90</p> <p>HMM build command line: hmmbuild HMM SEED</p> <p>HMM build command line: hmmcalibrate --seed 0 HMM</p> <p>Database Reference INTERPRO; IPR002674;</p> <p>Comment: This ribosomal protein is found in archaeobacteria and eukaryotes. It contains four conserved cysteine residues that may bind to zinc.</p> <p>Number of members: 15</p>
Ribosomal_L37e	PDOC00827	Ribosomal protein L37e signature	<p>A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:</p> <ul style="list-style-type: none"> - Mammalian L37 [1]. - Leishmania infantum L37 [2]. - Fission yeast YL35 [3]. - Halobacterium marismortui L37e (L35e) [4]. <p>These proteins have 56 to 96 amino-acid residues. As a signature pattern, we</p>

00689980 101300

Pfam	Prosite	Full Name	Description
			<p>HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 88230452 Reference Title: Interaction of proteins S16, S17 and S20 with 16 S Reference Title: ribosomal RNA. Reference Author: Stern S, Changchien LM, Craven GR, Noller HF; Reference Location: J Mol Biol 1988;200:291-299. Database Reference INTERPRO; IPR002583; Comment: Bacterial ribosomal protein S20 interacts with 16S rRNA [1]. Number of members: 29</p>
Ribosomal_S27e	PDOC00898	Ribosomal protein S27e signature	<p>A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of [1]:</p> <ul style="list-style-type: none"> - Mammalian S27 (human S27 was originally known as metallopeptide-stimulin 1). - Chlamydomonas reinhardtii S27. - Entamoeba histolytica S27. - Yeast S27. - Archaeobacterial S27e. <p>These proteins have from 62 to 87 amino acids. They contain, in their central section, a putative zinc-finger region of the type C-x(2)-C-x(14)-C-x(2)-C. We have selected that region as a signature pattern.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [QKT]-C-x(2)-C-x(6)-F-[GSD]-x-[PSA]-x(5)-C-x(2)-C-[GSA]-x(2)-[LV]-x(2)-P-x-G [The four C's are potential zinc ligands] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update December 1999 / Pattern and text revised. References [1] Chan Y.-L., Suzuki K., Olvera J., Wool I.G. Nucleic Acids Res. 21:649-655(1993).</p>
Ribosomal_S3_C	PDOC00474	Ribosomal protein S3 signature	<p>Ribosomal protein S3 is one of the proteins from the small ribosomal subunit. In Escherichia coli, S3 is known to be involved in the binding of initiator Met-tRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups:</p> <ul style="list-style-type: none"> - Eubacterial S3. - Algal and plant chloroplast S3. - Cyanelle S3. - Archaeobacterial S3. - Plant mitochondrial S3. - Vertebrate S3. - Insect S3. - Caenorhabditis elegans S3 (C23G10.3). - Yeast S3 (Rp13). <p>S3 is a protein of 209 to 559 amino-acid residues. As signature patterns, we selected a conserved region located in the C-terminal section.</p>

09689980 " 101300

Pfam	Prosite	Full Name	Description
			<p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [GSTA]-[KR]-x(6)-G-x-[LIVMT]-x(2)-[NQSCH]-x(1,3)-[LIVFCA]-x(3)-[LIV]-[DENQ]-x(7)-[LMT]-x(2)-G-x(2)-[GS] Sequences known to belong to this class detected by the pattern ALL, except for some mitochondrial S3. Other sequence(s) detected in SWISS-PROT NONE. Expert(s) to contact by email Hallick R.B. hallick@arizona.edu</p> <p>Last update December 1999 / Pattern and text revised. References [1] Otaka E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).</p>
Ribosomal_S3_N	PDOC00474	Ribosomal protein S3 signature	<p>Ribosomal protein S3 is one of the proteins from the small ribosomal subunit. In Escherichia coli, S3 is known to be involved in the binding of initiator Met-tRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups:</p> <ul style="list-style-type: none"> - Eubacterial S3. - Algal and plant chloroplast S3. - Cyanelle S3. - Archaeobacterial S3. - Plant mitochondrial S3. - Vertebrate S3. - Insect S3. - Caenorhabditis elegans S3 (C23G10.3). - Yeast S3 (Rp13). <p>S3 is a protein of 209 to 559 amino-acid residues. As signature patterns, we selected a conserved region located in the C-terminal section.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [GSTA]-[KR]-x(6)-G-x-[LIVMT]-x(2)-[NQSCH]-x(1,3)-[LIVFCA]-x(3)-[LIV]-[DENQ]-x(7)-[LMT]-x(2)-G-x(2)-[GS] Sequences known to belong to this class detected by the pattern ALL, except for some mitochondrial S3. Other sequence(s) detected in SWISS-PROT NONE. Expert(s) to contact by email Hallick R.B. hallick@arizona.edu</p> <p>Last update December 1999 / Pattern and text revised. References [1] Otaka E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).</p>
RimM		RimM	<p>Accession number: PF01782 Definition: RimM Author: Bateman A Alignment method of seed: Clustalw Source of seed members: PSI-BLAST P51419 Gathering cutoffs: 25 25 Trusted cutoffs: 49.00 49.00 Noise cutoffs: -66.10 -66.10 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98083058 Reference Title: RimM and RbfA are essential for efficient processing of 16S Reference Title: rRNA in Escherichia coli. Reference Author: Bylund GO, Wipemo LC, Lundberg LA,</p>

00589980 101300

Pfam	Prosite	Full Name	Description
			<p>Wikstrom PM; Reference Location: J Bacteriol 1998;180:73-82. Database Reference INTERPRO; IPR002676; Comment: The RimM protein is essential for efficient processing of 16S rRNA [1]. Comment: The RimM protein was shown to have affinity for free ribosomal 30S subunits but not for 30S subunits in the 70S ribosomes [1]. Number of members: 14</p>
RNA_dep_RNA_pol		RNA dependent RNA polymerase	<p>Accession number: PF00680 Definition: RNA dependent RNA polymerase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_32 (release 2.1) Gathering cutoffs: -127 -127 Trusted cutoffs: -117.00 -117.00 Noise cutoffs: -137.30 -137.30 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Database Reference: SCOP; 1rdr; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR001205; Database Reference PDB; 1rdr ; 12; 37; Database Reference PDB; 1rdr ; 182; 460; Database Reference PDB; 1rdr ; 67; 97; Database reference: PFAMB; PB039844; Database reference: PFAMB; PB040630; Database reference: PFAMB; PB040631; Database reference: PFAMB; PB040844; Database reference: PFAMB; PB041022; Database reference: PFAMB; PB041498; Number of members: 271</p>
RNA_dep_RNApol2		RNA dependent RNA polymerase	<p>Accession number: PF00978 Definition: RNA dependent RNA polymerase Author: Finn RD, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_13 (release 3.0) Gathering cutoffs: 8.5 0 Trusted cutoffs: 8.50 0.20 Noise cutoffs: 8.40 8.40 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 93188140 Reference Title: Roles of nonstructural polyproteins and cleavage products in regulating Sindbis virus RNA replication and transcription. Reference Author: Lemm JA, Rice CM; Reference Location: J Virol 1993;67:1916-1926. Reference Number: [2] Reference Medline: 96323143 Reference Title: Complete replication in vitro of tobacco mosaic virus RNA by a template-dependent, membrane-bound RNA polymerase. Reference Author: Osman TA, Buck KW; Reference Location: J Virol 1996;70:6227-6234. Reference Number: [3] Reference Medline: 94047331 Reference Title: Bromovirus RNA replication and transcription require compatibility between the polymerase- and helicase-like viral RNA synthesis proteins. Reference Author: Dinant S, Janda M, Kroner PA, Ahlquist P; Reference Location: J Virol 1993;67:7181-7189. Reference Number: [4] Reference Medline: 94094568 Reference Title: Evolution and taxonomy of positive-strand</p>

0968980 10300

Pfam	Prosite	Full Name	Description
			<p>including mitochondrial. Comment: and chloroplast polymerases). Comment: This family includes a region of about 400 amino acids. Comment: This family includes the whole archaeobacterial A" subunit, Comment: but only the C terminal region of the A subunit from eukaryotes Comment: and the beta' subunit from eubacteria. Number of members: 105</p>
RNB	PDOC00904	Ribonuclease II family signature	<p>On the basis of sequence similarities, the following bacterial and eukaryotic proteins seem to form a family:</p> <ul style="list-style-type: none"> - Escherichia coli and related bacteria ribonuclease II (EC 3.1.13.1) (RNase II) (gene rnb) [1]. RNase II is an exonuclease involved in mRNA decay. It degrades mRNA by hydrolyzing single-stranded polyribonucleotides processively in the 3' to 5' direction. - Bacterial ribonuclease R [2], a 3'-5'exoribonuclease that participates in an essential cell function. - Yeast protein SSD1 (or SRK1) which is implicated in the control of the cell cycle G1 phase. - Yeast protein DIS3 [3], which binds to ran (GSP1) and enhances the nucleotide-releasing activity of RCC1 on ran. - Fission yeast protein dis3, which is implicated in mitotic control. - Neurospora crassa cyt-4, a mitochondrial protein required for RNA 5' and 3' end processing and splicing. - Yeast protein MSU1, which is involved in mitochondrial biogenesis. - Synechocystis strain PCC 6803 protein zam [4], which control resistance to the carbonic anhydrase inhibitor acetazolamide. - Caenorhabditis elegans hypothetical protein F48E8.6. <p>The size of these proteins range from 644 residues (rnb) to 1250 (SSD1). While their sequence is highly divergent they share a conserved domain in their C-terminal section [5]. It is possible that this domain plays a role in a putative exonuclease function that would be common to all these proteins. We have developed a signature pattern based on the core of this conserved domain.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [HI]-[FYE]-[GSTAM]-[LIVM]-x(4,5)-Y-[STALV]-x-[FWVAC]-[TV]-[SA]-P-[LIVMA]-[RQ]-[KR]-[FY]-x-D-x(3)-[HQ] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update December 1999 / Pattern and text revised. References [1] Zilhao R., Camelo L., Arraiano C.M. Mol. Microbiol. 8:43-51(1993). [2] Cheng Z.-F., Zuo Y., Li Z., Rudd K.E., Deutscher M.P. J. Biol. Chem. 273:14077-14080(1998). [3]</p>

05689980-10100

Pfam	Prosite	Full Name	Description
			Database Reference PDB; 1vsm A; 54; 199;
			Database Reference PDB; 1czb A; 53; 198;
			Database Reference PDB; 1asw ; 53; 201;
			Database Reference PDB; 1cz9 A; 59; 197;
			Database Reference PDB; 1vsk ; 54; 199;
			Database Reference PDB; 1vsl A; 54; 199;
			Database Reference PDB; 1asu ; 53; 207;
			Database Reference PDB; 1c0m A; 53; 213;
			Database Reference PDB; 1vsd ; 54; 88;
			Database Reference PDB; 1vse ; 54; 88;
			Database Reference PDB; 1c1a B; 55; 213;
			Database Reference PDB; 1c0m B; 54; 213;
			Database Reference PDB; 1c0m D; 54; 213;
			Database Reference PDB; 1c1a A; 53; 213;
			Database Reference PDB; 1c0m C; 53; 213;
			Database Reference PDB; 1bhl ; 57; 201;
			Database Reference PDB; 1bi4 B; 57; 201;
			Database Reference PDB; 1bl3 B; 57; 201;
			Database Reference PDB; 1b9f A; 56; 201;
			Database Reference PDB; 1bis B; 56; 201;
			Database Reference PDB; 1qs4 B; 56; 201;
			Database Reference PDB; 1qs4 C; 56; 201;
			Database Reference PDB; 1biz A; 54; 201;
			Database Reference PDB; 1itg ; 55; 201;
			Database Reference PDB; 1bi4 C; 53; 201;
			Database Reference PDB; 1bl3 C; 53; 201;
			Database Reference PDB; 2itg ; 53; 201;
			Database Reference PDB; 1b9d A; 57; 189;
			Database Reference PDB; 1bi4 A; 57; 201;
			Database Reference PDB; 1bl3 A; 57; 201;
			Database Reference PDB; 1bis A; 56; 201;
			Database Reference PDB; 1biu A; 56; 201;
			Database Reference PDB; 1biu B; 56; 201;
			Database Reference PDB; 1biu C; 56; 201;
			Database Reference PDB; 1qs4 A; 56; 201;
			Database Reference PDB; 1b92 A; 56; 201;
			Database Reference PDB; 1biz B; 58; 201;
			Database Reference PDB; 1b9d A; 382; 390;
			Database Reference PDB; 1wjb A; 53; 55;
			Database Reference PDB; 1wjb B; 53; 55;
			Database Reference PDB; 1wjd A; 53; 55;
			Database Reference PDB; 1wjd B; 53; 55;
			Database Reference PDB; 1wjf A; 53; 55;
			Database Reference PDB; 1wjf B; 53; 55;
			Database reference: PFAMB; PB000048;
			Database reference: PFAMB; PB007709;
			Database reference: PFAMB; PB013923;
			Database reference: PFAMB; PB013938;
			Database reference: PFAMB; PB018509;
			Database reference: PFAMB; PB020302;
			Database reference: PFAMB; PB025327;
			Database reference: PFAMB; PB028352;
			Database reference: PFAMB; PB032740;
			Database reference: PFAMB; PB040612;
			Database reference: PFAMB; PB040636;
			Database reference: PFAMB; PB040684;
			Database reference: PFAMB; PB040695;
			Database reference: PFAMB; PB040730;
			Database reference: PFAMB; PB040824;
			Database reference: PFAMB; PB041112;
			Database reference: PFAMB; PB041143;
			Database reference: PFAMB; PB041275;
			Database reference: PFAMB; PB041356;
			Database reference: PFAMB; PB041375;
			Database reference: PFAMB; PB041456;
			Database reference: PFAMB; PB041459;
			Database reference: PFAMB; PB041522;
			Database reference: PFAMB; PB041665;
			Database reference: PFAMB; PB041761;
			Database reference: PFAMB; PB041816;
			Database reference: PFAMB; PB041885;
			Comment: Integrase mediates integration of a DNA
			copy of the viral
			Comment: genome into the host chromosome.

09689980 101300

Pfam	Prosite	Full Name	Description
			<p>Integrase is composed of</p> <p>Comment: three domains. The amino-terminal domain is a zinc binding</p> <p>Comment: domain Integrase_Zn. This domain is the central catalytic</p> <p>Comment: domain. The carboxyl terminal domain that is a non-specific</p> <p>Comment: DNA binding domain integrase.</p> <p>Comment: The catalytic domain acts as an endonuclease when two</p> <p>Comment: nucleotides are removed from the 3' ends of the blunt-ended</p> <p>Comment: viral DNA made by reverse transcription.</p> <p>Comment: This domain also</p> <p>Comment: catalyses the DNA strand transfer reaction of the 3' ends</p> <p>Comment: of the viral DNA to the 5' ends of the integration site [1].</p> <p>Number of members: 1147</p>
S4		S4 domain	<p>Accession number: PF01479</p> <p>Definition: S4 domain</p> <p>Author: Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Medline:99193178</p> <p>Gathering cutoffs: 17 17</p> <p>Trusted cutoffs: 17.20 17.20</p> <p>Noise cutoffs: 16.70 16.70</p> <p>HMM build command line: hmmbuild HMM SEED</p> <p>HMM build command line: hmmscalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 99193178</p> <p>Reference Title: Novel predicted RNA-binding domains associated with the</p> <p>Reference Title: translation machinery.</p> <p>Reference Author: Aravind L, Koonin EV;</p> <p>Reference Location: J Mol Evol 1999;48:291-302.</p> <p>Reference Number: [2]</p> <p>Reference Medline: 98372721</p> <p>Reference Title: The crystal structure of ribosomal protein S4 reveals a</p> <p>Reference Title: two-domain molecule with an extensive RNA-binding surface:</p> <p>Reference Title: one domain shows structural homology to the ETS DNA-binding</p> <p>Reference Title: motif.</p> <p>Reference Author: Davies C, Gerstner RB, Draper DE, Ramakrishnan V, White SW;</p> <p>Reference Location: EMBO J 1998;17:4545-4558.</p> <p>Database Reference: SCOP; 1c06; fa; [SCOP-USA][CATH-PDBSUM]</p> <p>Database Reference: INTERPRO; IPR002942;</p> <p>Database Reference: PDB; 1c05 A; 51; 98;</p> <p>Database Reference: PDB; 1c06 A; 51; 98;</p> <p>Database Reference: PDB; 1dm9 A; 9; 55;</p> <p>Database Reference: PDB; 1dm9 B; 9; 55;</p> <p>Database reference: PFAMB; PB001751;</p> <p>Database reference: PFAMB; PB041147;</p> <p>Database reference: PFAMB; PB041148;</p> <p>Comment: The S4 domain is a small domain consisting of 60-65 amino acid residues</p> <p>Comment: that was detected in the bacterial ribosomal protein S4, eukaryotic</p> <p>Comment: ribosomal S9, two families of pseudouridine synthases, a novel family</p> <p>Comment: of predicted RNA methylases, a yeast protein containing a pseudouridine</p> <p>Comment: synthetase and a deaminase domain, bacterial tyrosyl-tRNA synthetases,</p> <p>Comment: and a number of uncharacterized, small proteins that may be involved in</p> <p>Comment: translation regulation [1]. The S4 domain probably mediates binding to</p> <p>Comment: RNA.</p>

Pfam	Prosite	Full Name	Description
			Number of members: 256
SAA_proteins	PDOC00762	Serum amyloid A proteins signature	<p>The serum amyloid A (SAA) proteins comprise a family of vertebrate proteins that associate predominantly with high density lipoproteins (HDL) [1,2]. The synthesis of certain members of the family is greatly increased (as much as a 1000 fold) in inflammation; thus making SAA a major acute phase reactant. While the major physiological function of SAA is unclear, prolonged elevation of plasma SAA levels, as in chronic inflammation, however, results in a pathological condition, called amyloidosis, which affects the liver, kidney and spleen and which is characterized by the highly insoluble accumulation of SAA in these tissues.</p> <p>SAA are proteins of about 110 amino acid residues. As a signature pattern, we selected the most highly conserved region, which is located in the central part of the sequence.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern A-R-G-N-Y-[ED]-A-x-[QKR]-R-G-x-G-G-x-W-A Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Last update June 1994 / First entry. References [1] Malle E., Steinmetz A., Raynes J.G. Atherosclerosis 102:131-146(1993).</p> <p>[2] Uhlir C.M., Burgess C.J., Sharp P.M., Whitehead A.S. Genomics 19:228-235(1994).</p>
SAM		SAM domain (Sterile alpha motif)	<p>Accession number: PF00536 Definition: SAM domain (Sterile alpha motif) Author: Bateman A Alignment method of seed: Clustalw Source of seed members: [1],[2] Gathering cutoffs: 11 0 Trusted cutoffs: 11.00 3.70 Noise cutoffs: 10.90 10.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96100659 Reference Title: SAM: A novel motif in yeast sterile alpha and Drosophila Reference Title: polyhomeotic proteins Reference Author: Ponting CP; Reference Location: Prot Sci 1995;4:1928-1930. Reference Number: [2] Reference Medline: 97160498 Reference Title: SAM as a protein interaction domain involved in Reference Title: developmental regulation. Reference Author: Shultz J, Ponting CP, Hofmann K, Bork P; Reference Location: Prot Sci 1997;6:249-253. Reference Number: [3] Reference Medline: 99101382 Reference Title: The crystal structure of an Eph receptor SAM domain reveals Reference Title: a mechanism for modular dimerization.</p>

Pfam	Prosite	Full Name	Description
			<p>Reference Author: Stapleton D, Balan I, Pawson T, Sicheri F; Reference Location: Nat Struct Biol 1999;6:44-49. Database reference: SMART; SAM; Database Reference: SCOP; 1b0x; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR001660; Database Reference PDB; 1b0x A; 910; 973; Database Reference PDB; 1sgg ; 7; 70; Database Reference PDB; 1b4f A; 7; 71; Database Reference PDB; 1b4f C; 7; 71; Database Reference PDB; 1b4f E; 7; 71; Database Reference PDB; 1b4f D; 7; 71; Database Reference PDB; 1b4f H; 7; 71; Database Reference PDB; 1b4f F; 7; 71; Database Reference PDB; 1b4f G; 7; 71; Database Reference PDB; 1b4f B; 7; 71; Database reference: PFAMB; PB008631; Database reference: PFAMB; PB040678; Database reference: PFAMB; PB041111; Database reference: PFAMB; PB041385; Comment: It has been suggested that SAM is an evolutionarily conserved protein Comment: binding domain that is involved in the regulation of numerous Comment: developmental processes in diverse eukaryotes. Comment: The SAM domain can potentially function as a protein interaction Comment: module through its ability to homo- and heterooligomerise with Comment: other SAM domains. Number of members: 110</p>
SAM_decarbox		Adenosylmethionine decarboxylase	<p>Accession number: PF01536 Definition: Adenosylmethionine decarboxylase Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_600 (release 4.0) Gathering cutoffs: 11 11 Trusted cutoffs: 17.90 17.90 Noise cutoffs: 5.70 5.70 HMM build command line: hmmbuild -f HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 98098079 Reference Title: Cloning, mapping and mutational analysis of the Reference Title: S-adenosylmethionine decarboxylase gene in Drosophila Reference Title: melanogaster. Reference Author: Larsson J, Rasmuson-Lestander A; Reference Location: Mol Gen Genet 1997;256:652-660. Database Reference: SCOP; 1jen; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR001985; Database Reference PDB; 1jen C; 69; 328; Database Reference PDB; 1jen A; 69; 329; Database Reference PDB; 1jen B; 4; 67; Database Reference PDB; 1jen D; 5; 66; Comment: This is a family of S-adenosylmethionine decarboxylase (SAMDC) proenzymes. Comment: In the biosynthesis of polyamines SAMDC produces decarboxylated Comment: S-adenosylmethionine, which serves as the aminopropyl moiety necessary Comment: for spermidine and spermine biosynthesis from putrescine [1]. The Pfam Comment: alignment contains both the alpha and beta chains that are cleaved to Comment: form the active enzyme. Number of members: 34</p>
SBF		Sodium Bile acid symporter family	<p>Accession number: PF01758 Definition: Sodium Bile acid symporter family</p>

Pfam	Prosite	Full Name	Description
			<p>Gathering cutoffs: -50 -50 Trusted cutoffs: -48.00 -48.00 Noise cutoffs: -82.00 -82.00 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96048023 Reference Title: The molecular biology of multidomain proteins. Selected examples. Reference Author: Hawkins AR, Lamb HK; Reference Location: Eur J Biochem 1995;232:7-18. Database Reference: INTERPRO; IPR002907; Comment: This family contains both shikimate and quinate dehydrogenases. Comment: Shikimate 5-dehydrogenase catalyses the conversion of shikimate to 5-dehydroshikimate. This reaction is part of the shikimate pathway which is involved in the biosynthesis of aromatic amino acids. Comment: Quinate 5-dehydrogenase catalyses the conversion of quinate to 5-dehydroquininate. This reaction is part of the quinate pathway where quinic acid is exploited as a source of carbon in prokaryotes and eukaryotes. Comment: Both the shikimate and quinate pathways share two common pathway metabolites 3-dehydroquininate and dehydroshikimate. Number of members: 58</p>
Sigma54_factors	PDOC00593	Sigma-54 factors family signatures and profile	<p>Sigma factors [1] are bacterial transcription initiation factors that promote the attachment of the core RNA polymerase to specific initiation sites and are then released. They alter the specificity of promoter recognition. Most bacteria express a multiplicity of sigma factors. Two of these factors, sigma-70 (gene rpoD), generally known as the major or primary sigma factor, and sigma-54 (gene rpoN or ntrA) direct the transcription of a wide variety of genes. The other sigma factors, known as alternative sigma factors, are required for the transcription of specific subsets of genes.</p> <p>With regard to sequence similarity, sigma factors can be grouped into two classes: the sigma-54 and sigma-70 families. The sigma-70 family has many different sigma factors (see the relevant entry <PDOC00592>). The sigma-54 family consists exclusively of sigma-54 factor [2,3] required for the transcription of promoters that have a characteristic -24 and -12 consensus recognition element but which are devoid of the typical -10,-35 sequences recognized by the major sigma factors. The sigma-54 factor is also characterized by its interaction with ATP-dependent positive regulatory proteins that bind to upstream activating sequences.</p> <p>Structurally sigma-54 factors consist of three distinct regions:</p> <ul style="list-style-type: none"> - A relatively well conserved N-terminal glutamine-rich region of

Pfam	Prosite	Full Name	Description
			<p>about 50 residues that contains a potential leucine zipper motif.</p> <ul style="list-style-type: none"> - A region of variable length which is not well conserved. - A well conserved C-terminal region of about 350 residues that contains a second potential leucine zipper, a potential DNA-binding 'helix-turn-helix' motif and a perfectly conserved octapeptide whose function is not known. <p>We developed two signature patterns for this family of sigma factors. The first starts two residues before the N-terminal extremity of the helix-turn-helix region and ends two residues before its C-terminal extremity. The second is the conserved octapeptide. A profile has also been designed that covers the whole C-terminal region.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern P-[LIVM]-x-[LIVM]-x(2)-[LIVM]-A-x(2)-[LIVMFT]-x(2)-[HS]-x- S-T-[LIVM]-S-R Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Consensus pattern R-R-T-[IV]-[ATN]-K-Y-R Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so.</p> <p>Last update July 1999 / Patterns and text revised.</p> <p>References [1] Helmann J.D., Chamberlin M.J. Annu. Rev. Biochem. 57:839-872(1988).</p> <p>[2] Thoeny B., Hennecke H. FEMS Microbiol. Rev. 5:341-358(1989).</p> <p>[3] Merrick M.J. Mol. Microbiol. 10:903-909(1993).</p>
SLH	PDOC00823	S-layer homology domain signature	<p>S-layers are paracrystalline mono-layered assemblies of (glyco)proteins which coat the surface of bacteria [1]. Several S-layer proteins and some other cell wall proteins contain one or more copies of a domain of about 50-60 residues, which has been called SLH (for S-layer homology) [2]. There is strong evidence that this domain serves as an anchor to the peptidoglycan [3]. The SLH domain has been found in:</p> <ul style="list-style-type: none"> - S-layer glycoprotein of <i>Acetogenium kivui</i> (3 copies). - S-layer 125 Kd protein of <i>Bacillus sphaericus</i> (3 copies). - S-layer protein of <i>Bacillus anthracis</i> (3 copies). - S-layer protein of <i>Bacillus licheniformis</i> (3 copies).

09689980 101300

Pfam	Prosite	Full Name	Description
			<p>- S-layer protein (HWP) from <i>Bacillus brevis</i> strain HPD31 (3 copies).</p> <p>- Middle cell wall protein (MWP) from <i>Bacillus brevis</i> strain 47 (3 copies).</p> <p>- S-layer protein (p100) of <i>Thermus thermophilus</i> (1 copy).</p> <p>- Outer membrane protein Omp-alpha from <i>Thermotoga maritima</i> (1 copy).</p> <p>- Cellulosome anchoring protein (gene ancA), outer layer protein B (OlpB) and a further potential cell surface glycoprotein from <i>Clostridium thermocellum</i> (3 copies; the first copy is missing its N-terminal third which is appended to the end of the third copy; may have arisen by circular permutation).</p> <p>- Amylopullulanase (gene amyB) from <i>Thermoanaerobacter thermosulfurogenes</i> (3 copies)</p> <p>- Amylopullulanase (gene aapT) from <i>Bacillus</i> strain XAL-601 (3 copies).</p> <p>- Endoglucanase from <i>Bacillus</i> strain KSM-635 (3 copies).</p> <p>- Exoglucanase (gene xynX) from <i>Clostridium thermocellum</i> (3 copies).</p> <p>- Xylanase A (gene xynA) from <i>Thermoanaerobacter saccharolyticum</i> (2 copies; 3 copies if a frameshift is taken into account).</p> <p>- Protein involved in butirosin production (ButB) from <i>Bacillus circulans</i> (2 incomplete copies; 3 copies if three frameshifts are taken into account).</p> <p>- Two hypothetical proteins from <i>Synechocystis</i> strain PCC 6803 (1 copy each).</p> <p>- A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from <i>Bacillus circulans</i> (fragment of 1 copy; 3 copies if two frameshifts are taken into account).</p> <p>SLH domains are found at the N- or C-termini of mature proteins. They occur in single copy followed by a predicted coiled coil domain, or in three contiguous copies. Structurally, the SLH domain is predicted to contain two alpha-helices flanking a beta strand. The SLH sequences are fairly divergent with an average identity of about 25%. It is however possible to build a sequence pattern that starts at the second position of the domain and that spans 3/4 of its length.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [LVFYT]-x-[DA]-x(2,5)-[DNQSATPHY]-[FYWPDA]-x(4)-[LIV]-x(2)-[GTALV]-x(4,6)-[LIVFYC]-x(2)-G-x-[PGSTA]-x(2,3)-[MFYA]-x-[PGAV]-x(3,10)-[LIVMA]-[STKR]-[RY]-x-[EQ]-x-[STALIVM]</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Expert(s) to contact by email Lupas A.N. lupas@vms.biochem.mpg.de</p> <p>Last update November 1997 / Pattern and text revised.</p> <p>References [1] Beveridge T.J. Curr. Opin. Struct. Biol. 4:204-212(1994).</p> <p>[2] Lupas A., Engelhardt H., Peters J., Santarius U., Volker S.,</p>

Pfam	Prosite	Full Name	Description
			<p>Baumeister W. J. Bacteriol. 176:1224-1233(1994).</p> <p>[3] Lemaire M., Ohayon H., Gounon P., Fujino T., Beguin P. J. Bacteriol. 177:2451-2459(1995).</p>
Smr		Smr domain	<p>Accession number: PF01713 Definition: Smr domain Author: Bateman A Alignment method of seed: Clustalw Source of seed members: [1] Gathering cutoffs: 0 0 Trusted cutoffs: 1.40 1.40 Noise cutoffs: -7.90 -7.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 10431172 Reference Title: Smr: a bacterial and eukaryotic homologue of the C-terminal region of the MutS2 family. Reference Author: Moreira D, Philippe H; Reference Location: Trends Biochem Sci 1999;24:298-300. Database Reference: INTERPRO; IPR002625; Comment: This family includes the Smr (Small MutS Related) proteins, Comment: and the C-terminal region of the MutS2 protein. It has been Comment: suggested that this domain interacts with the MutS1 Comment: Swiss:P23909 protein in the case of Smr proteins and with Comment: the N-terminal MutS related region of MutS2 Swiss:P94545 [1]. Number of members: 14</p>
SRF-TF	PDOC00302	MADS-box domain signature and profile	<p>A number of transcription factors contain a conserved domain of 56 amino-acid residues, sometimes known as the MADS-box domain [E1]. They are listed below:</p> <ul style="list-style-type: none"> - Serum response factor (SRF) [1], a mammalian transcription factor that binds to the Serum Response Element (SRE). This is a short sequence of dyad symmetry located 300 bp to the 5' end of the transcription initiation site of genes such as c-fos. - Mammalian myocyte-specific enhancer factors 2A to 2D (MEF2A to MEF2D). These proteins are transcription factor which binds specifically to the MEF2 element present in the regulatory regions of many muscle-specific genes. - Drosophila myocyte-specific enhancer factor 2 (MEF2). - Yeast GRM/PRTF protein (gene MCM1) [2], a transcriptional regulator of mating-type-specific genes. - Yeast arginine metabolism regulation protein I (gene ARGR1 or ARG80). - Yeast transcription factor RLM1. - Yeast transcription factor SMP1. - Arabidopsis thaliana agamous protein (AG) [3], a probable transcription factor involved in regulating genes that determines stamen and carpel development in wild-type flowers. Mutations in the AG gene result in the replacement of the stamens by petals and the carpels by a new flower. - Arabidopsis thaliana homeotic proteins Apetala1 (AP1), Apetala3 (AP3) and

"03669980" 03669980

Pfam	Prosite	Full Name	Description
			<p>- Drosophila MtSSB. - Yeast protein RIM1.</p> <p>We have developed two signature patterns for these proteins. The first is a conserved region in the N-terminal section of the SSB's. The second is a centrally located region which, in Escherichia coli SSB, is known to be involved in the binding of DNA.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [LIVMF]-[NST]-[KRHST]-[LIVM]-x-[LIVMF](2)-G-[NHRK]-[LIVMA]-[GST]-x-[DENT] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Consensus pattern T-x-W-[HY]-[RNS]-[LIVM]-x-[LIVMF]-[FY]-[NGKR] Sequences known to belong to this class detected by the pattern A majority. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update December 1999 / Patterns and text revised.</p> <p>References [1] Meyer R.R., Laine P.S. Microbiol. Rev. 54:342-380(1990).</p> <p>[2] Stroumbakis N.D., Li Z., Tolia P.P. Gene 143:171-177(1994).</p>
START		START domain	<p>Accession number: PF01852 Definition: START domain Author: SMART Alignment method of seed: Manual Source of seed members: Alignment kindly provided by SMART Gathering cutoffs: 25 25 Trusted cutoffs: 106.20 106.20 Noise cutoffs: -20.90 -20.90 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 99257451 Reference Title: START: a lipid-binding domain in StAR, HD-ZIP and Reference Title: signalling proteins. Reference Author: Ponting CP, Aravind L; Reference Location: Trends Biochem Sci 1999;24:130-132. Database reference: SMART; START; Database Reference INTERPRO; IPR002913; Number of members: 41</p>
Sterol_desat		Sterol desaturase	<p>Accession number: PF01598 Definition: Sterol desaturase Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_905 (release 4.1) Gathering cutoffs: -13 -13 Trusted cutoffs: 12.90 12.90 Noise cutoffs: -44.50 -44.50 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmcalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 91323727 Reference Title: Cloning, disruption and sequence of the gene encoding yeast Reference Title: C-5 sterol desaturase. Reference Author: Arthington BA, Bennett LG, Skatrud PL,</p>

09689980 1009

Pfam	Prosite	Full Name	Description
			<p>Last update November 1997 / Pattern and text revised.</p> <p>References</p> <p>[1] Benkovic S.J. Annu. Rev. Biochem. 49:227-251(1980).</p> <p>[2] Ross P., O'Gara F., Condon S. Appl. Environ. Microbiol. 56:2156-2163(1990).</p>
Top6A		Type II DNA topoisomerase	<p>Accession number: PF01962</p> <p>Definition: Type II DNA topoisomerase</p> <p>Author: Enright A, Ouzounis C, Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Enright A</p> <p>Gathering cutoffs: -99 -99</p> <p>Trusted cutoffs: -40.40 -40.40</p> <p>Noise cutoffs: -158.40 -158.40</p> <p>HMM build command line: hmmbuild -F HMM SEED</p> <p>HMM build command line: hmmcalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 97238688</p> <p>Reference Title: An atypical topoisomerase II from Archaea with implications for meiotic recombination [see comments]</p> <p>Reference Author: Bergerat A, de Massy B, Gabelle D, Varoutas PC, Nicolas A,</p> <p>Reference Author: Forterre P;</p> <p>Reference Location: Nature 1997;386:414-417.</p> <p>Database Reference: SCOP; 1d3y; fa; [SCOP-USA][CATH-PDBSUM]</p> <p>Database Reference INTERPRO; IPR002815;</p> <p>Database Reference PDB; 1d3y A; 77; 363;</p> <p>Database Reference PDB; 1d3y B; 77; 363;</p> <p>Comment: Members of this family are the A subunit from type II DNA</p> <p>Comment: topoisomerases. Type II DNA</p> <p>Comment: topoisomerases catalyse the relaxation</p> <p>Comment: of DNA supercoiling by causing transient double strand breaks.</p> <p>Comment: The family includes topoisomerase VI subunit A from archaeobacteria</p> <p>Comment: Swiss:Q57815 EC:5.99.1.3 and SPO11 from yeast Swiss:P23179.</p> <p>Comment: A conserved tyrosine is thought to be involved in breaking the</p> <p>Comment: double stranded DNA [1].</p> <p>Number of members: 9</p>
Topoisom_bac	PDOC00333	Prokaryotic DNA topoisomerase I active site	<p>DNA topoisomerase I (EC 5.99.1.2) [1,2,3,4,E1] is one of the two types of enzyme that catalyze the interconversion of topological DNA isomers. Type I topoisomerases act by catalyzing the transient breakage of DNA, one strand at a time, and the subsequent rejoining of the strands. When a prokaryotic type I topoisomerase breaks a DNA backbone bond, it simultaneously forms a protein-DNA link where the hydroxyl group of a tyrosine residue is joined to a 5'-phosphate on DNA, at one end of the enzyme-severed DNA strand.</p> <p>Prokaryotic organisms, such as Escherichia coli, have two type I topoisomerase isozymes: topoisomerase I (gene topA) and topoisomerase III (gene topB).</p> <p>Eukaryotes also contain homologs of prokaryotic topoisomerase III.</p> <p>There are a number of conserved residues in the region around the active site</p>

09669960 101700

Pfam	Prosite	Full Name	Description
			<p>This family also includes</p> <p>Comment: Swiss:Q99598, that was found to interact with translin with yeast</p> <p>Comment: two-hybrid screen [1].</p> <p>Number of members: 10</p>
Transposase_19		Transposase 19	<p>Members of this family are capable of in vitro and/or in vivo insertion of a donor polynucleotide into a target polynucleotide. Such biological activity is useful for inserting DNA into host genome, for example, for cloning purposes to generate a desired vector in vitro.</p>
TRANSPOSASE_IS30	PDOC00801	Transposases, IS30 family, signature	<p>Autonomous mobile genetic elements such as transposon or insertion sequences (IS) encode an enzyme, called transposase, required for excising and inserting the mobile element. On the basis of sequence similarities, transposases can be grouped into various families. One of these families has been shown [1,2] to consist of transposases from the following elements:</p> <ul style="list-style-type: none"> - Is30 from <i>Escherichia coli</i>. - Is1086 from <i>Alcaligenes eutrophus</i>. - Is1161 from <i>Streptococcus salivarius</i>. - Is4351 (Tn4551) from <i>Bacteroides fragilis</i>. <p>These transposases are proteins of 340 to 380 amino acids. The best conserved region is located in their C-terminal section and is used as a signature pattern.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern R-G-x(2)-E-N-x-N-G-[LIVM](2)-R-[QE]-[LIVMFY](2)-P-K</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update November 1995 / First entry.</p> <p>References</p> <p>[1]</p> <p>Dong Q., Sadouk A., van der Lelie D., Taghavi S., Ferhat A., Nuyten J.M., Borremans B., Mergeay M., Toussaint A. J. Bacteriol. 174:8133-8138(1992).</p> <p>[2]</p> <p>Giffard P.M., Rathsam C., Kwan E., Kwan D.W.L., Bunny K.L., Koo S.-P., Jacques N.A. J. Gen. Microbiol. 139:913-920(1993).</p>
Transthyretin	PDOC00617	Transthyretin signatures	<p>Transthyretin (prealbumin) [1] is a thyroid hormone-binding protein that seems to transport thyroxine (T4) from the bloodstream to the brain. It is a protein of about 130 amino acids that assembles as a homotetramer and forms an internal channel that binds thyroxine. Transthyretin is mainly synthesized in the brain choroid plexus. In humans, variants of the protein are associated with distinct forms of amyloidosis.</p> <p>The sequence of transthyretin is highly conserved in vertebrates. A number of uncharacterized proteins also belong to this family:</p> <ul style="list-style-type: none"> - <i>Escherichia coli</i> hypothetical protein yedX. - <i>Bacillus subtilis</i> hypothetical protein yunM.

00589980-101300

Pfam	Prosite	Full Name	Description
			<p>bonded to a histidine, which itself is hydrogen-bonded to a serine. The sequences in the vicinity of the active site serine and histidine residues are well conserved in this family of proteases [1]. A partial list of proteases known to belong to the trypsin family is shown below.</p> <ul style="list-style-type: none"> - Acrosin. - Blood coagulation factors VII, IX, X, XI and XII, thrombin, plasminogen, and protein C. - Cathepsin G. - Chymotrypsins. - Complement components C1r, C1s, C2, and complement factors B, D and I. - Complement-activating component of RA-reactive factor. - Cytotoxic cell proteases (granzymes A to H). - Duodenase I. - Elastases 1, 2, 3A, 3B (protease E), leukocyte (medullasin). - Enterokinase (EC 3.4.21.9) (enteropeptidase). - Hepatocyte growth factor activator. - Hepsin. - Glandular (tissue) kallikreins (including EGF-binding protein types A, B, and C, NGF-gamma chain, gamma-renin, prostate specific antigen (PSA) and tonin). - Plasma kallikrein. - Mast cell proteases (MCP) 1 (chymase) to 8. - Myeloblastin (proteinase 3) (Wegener's autoantigen). - Plasminogen activators (urokinase-type, and tissue-type). - Trypsins I, II, III, and IV. - Trypsases. - Snake venom proteases such as ancrod, batroxobin, cerastobin, flavoxobin, and protein C activator. - Collagenase from common cattle grub and collagenolytic protease from Atlantic sand fiddler crab. - Apolipoprotein(a). - Blood fluke cercarial protease. - Drosophila trypsin like proteases: alpha, easter, snake-locus. - Drosophila protease stubble (gene sb). - Major mite fecal allergen Der p III. <p>All the above proteins belong to family S1 in the classification of peptidases [2,E1] and originate from eukaryotic species. It should be noted that bacterial proteases that belong to family S2A are similar enough in the regions of the active site residues that they can be picked up by the same patterns. These proteases are listed below.</p> <ul style="list-style-type: none"> - Achromobacter lyticus protease I. - Lysobacter alpha-lytic protease. - Streptogrisin A and B (Streptomyces proteases A and B). - Streptomyces griseus glutamyl endopeptidase II. - Streptomyces fradiae proteases 1 and 2. <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [LIVM]-[ST]-A-[STAG]-H-C [H is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for complement components C1r and C1s, pig plasminogen, bovine protein C, rodent urokinase, ancrod, gyroxin and two insect trypsins. Other sequence(s) detected in SWISS-PROT 14.</p>

Pfam	Prosite	Full Name	Description
			<p>consist of a</p> <p>N-terminal ubiquitin-like protein of 74 residues fused to ribosomal protein S30.</p> <ul style="list-style-type: none"> - Mouse protein NEDD-8 [6], a ubiquitin-like protein of 81 residues. - Human protein BAT3, a large fusion protein of 1132 residues that contains a <p>N-terminal ubiquitin-like domain.</p> <ul style="list-style-type: none"> - Caenorhabditis elegans protein ubl-1 [7]. Ubl-1 is a fusion protein which <p>consist of a N-terminal ubiquitin-like protein of 70 residues fused to</p> <p>ribosomal protein S27A.</p> <ul style="list-style-type: none"> - Yeast DNA repair protein RAD23 [8]. RAD23 contains a N-terminal domain that <p>seems to be distantly, yet significantly, related to ubiquitin.</p> <ul style="list-style-type: none"> - Mammalian RAD23-related proteins RAD23A and RAD23B. - Mammalian BCL-2 binding athanogene-1 (BAG-1). BAG-1 is a protein of 274 <p>residues that contains a central ubiquitin-like domain.</p> <ul style="list-style-type: none"> - Human spliceosome associated protein 114 (SAP 114 or SF3A120). - Yeast protein DSK2, a protein involved in spindle pole body duplication and <p>which contains a N-terminal ubiquitin-like domain.</p> <ul style="list-style-type: none"> - Human protein CKAP1/TFCB, Schizosaccharomyces pombe protein alp11 and <p>Caenorhabditis elegans hypothetical protein F53F4.3. These proteins contain</p> <p>a N-terminal ubiquitin domain and a C-terminal CAP-Gly domain (see</p> <p><PDOC00660>).</p> <ul style="list-style-type: none"> - Schizosaccharomyces pombe hypothetical protein SpAC26A3.16. This protein <p>contains a N-terminal ubiquitin domain.</p> <ul style="list-style-type: none"> - Yeast protein SMT3. - Human ubiquitin-like proteins SMT3A and SMT3B. - Human ubiquitin-like protein SMT3C (also known as PIC1; Ubl1, Sumo-1; Gmp-1 <p>or Sentrin). This protein is involved in targeting ranGAP1 to the nuclear</p> <p>pore complex protein ranBP2.</p> <ul style="list-style-type: none"> - SMT3-like proteins in plants and Caenorhabditis elegans. <p>To identify ubiquitin and related proteins we have developed a pattern based</p> <p>on conserved positions in the central section of the sequence. A profile was</p> <p>also developed that spans the complete length of the ubiquitin domain.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern K-x(2)-[LIVM]-x-[DESAK]-x(3)-[LIVM]-[PA]-x(3)-Q-x-[LIVM]-[LIVMC]-[LIVMFY]-x-G-x(4)-[DE]</p> <p>Sequences known to belong to this class detected by the pattern ALL, except for the RAD23 and SMT3 subfamilies, BAG-1 and SAP 114.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Sequences known to belong to this class detected by the profile ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so.</p> <p>Last update</p> <p>July 1998 / Text revised.</p>

00689980 101300

Pfam	Prosite	Full Name	Description
			<ul style="list-style-type: none"> - Escherichia coli hypothetical protein yjbQ. - Mycobacterium tuberculosis hypothetical protein MtCY9C4.12. - Synechocystis strain PCC 6803 hypothetical protein sli1880. - Archaeoglobus fulgidus hypothetical protein AF2050. - Methanococcus jannaschii hypothetical protein MJ1081. - Methanobacterium thermoautotrophicum hypothetical protein MTH771. - Fission yeast hypothetical protein SpAC4A8.02c. <p>These are small proteins of 14 to 16 Kd. They can be picked up in the database by the following pattern. This pattern is located in the C-terminal part of these proteins.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern S-X(2)-[LIV]-x-[LIV]-x(2)-G-x(4)-G-T-W-Q-x-[LIV]</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update July 1998 / First entry.</p> <p>References [1]</p> <p>Bairoch A.</p> <p>Unpublished observations (1998).</p>
UPF0052		Uncharacterised protein family UPF0052	<p>Accession number: PF01933</p> <p>Definition: Uncharacterised protein family UPF0052</p> <p>Author: Enright A, Ouzounis C, Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Enright A</p> <p>Gathering cutoffs: 25 25</p> <p>Trusted cutoffs: 263.90 263.90</p> <p>Noise cutoffs: -134.40 -134.40</p> <p>HMM build command line: hmmbuild -F HMM SEED</p> <p>HMM build command line: hmmcalibrate --seed 0 HMM</p> <p>Database Reference INTERPRO; IPR002882;</p> <p>Number of members: 12</p>
UPF0057	PDOC01013	Uncharacterized protein family UPF0057 signature	<p>The following uncharacterized proteins have been shown [1] to be evolutionary related:</p> <ul style="list-style-type: none"> - Barley low-temperature induced protein blt101. - Lophorium elongatum salt-sress induced protein ESI3. - Yeast hypothetical proteins YDL123w, YDR276c, YDR525Bw and YJL151c. - Caenorhabditis elegans hypothetical proteins F47B7.1, T23F2.3, T23F2.4, T23F2.5 and ZK632.10. - Escherichia coli hypothetical protein yqaE. - Synechocystis strain PCC 6803 hypothetical protein ssr1169. <p>These are small proteins of from 52 to 140 amino-acid residues that contains two transmembrane domains. As a signature pattern we selected a region that corresponds to the end of the first transmembrane helix.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [LIV]-x-[STA]-[LIVF](3)-P-P-[LIVA]-[GA]-[IV]-x(4)-[GKN]</p> <p>Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Last update</p>

09689980 101300

Pfam	Prosite	Full Name	Description
			<p>Trusted cutoffs: 186.00 186.00 Noise cutoffs: -42.60 -42.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 97352660 Reference Title: Characterization of UreG, identification of a Reference Title: UreD-UreF-UreG complex, and evidence suggesting that a Reference Title: nucleotide-binding site in UreG is required for in vivo Reference Title: metallocenter assembly of Klebsiella aerogenes urease. Reference Author: Moncrief MB, Hausinger RP; Reference Location: J Bacteriol 1997;179:4081-4086. Reference Number: [2] Reference Medline: 96146510 Reference Title: Organization of Ureaplasma urealyticum urease gene cluster Reference Title: and expression in a suppressor strain of Escherichia coli. Reference Author: Neyrolles O, Ferris S, Behbahani N, Montagnier L, Blanchard Reference Author: A; Reference Location: J Bacteriol 1996;178:647-655. Reference Number: [3] Reference Medline: 94211837 Reference Title: In vitro activation of urease apoprotein and role of UreD Reference Title: as a chaperone required for nickel metallocenter assembly. Reference Author: Park IS, Carr MB, Hausinger RP; Reference Location: Proc Natl Acad Sci U S A 1994;91:3233- 3237. Database Reference INTERPRO; IPR002669; Comment: UreD is a urease accessory protein. Urease urease hydrolyses Comment: urea into ammonia and carbamic acid [2]. UreD is involved in Comment: activation of the urease enzyme via the UreD-UreF-UreG-urease complex Comment: [1] and is required for urease nickel metallocenter assembly [3]. Comment: See also UreF UreF, UreG HypB UreG. Number of members: 23</p>
UreF		UreF	<p>Accession number: PF01730 Definition: UreF Author: Bashton M, Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_2037 (release 4.1) Gathering cutoffs: -31 -31 Trusted cutoffs: -14.30 -14.30 Noise cutoffs: -49.30 -49.30 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96404789 Reference Title: Purification and activation properties of UreD-UreF-urease Reference Title: apoprotein complexes. Reference Author: Moncrief MB, Hausinger RP; Reference Location: J Bacteriol 1996;178:5417-5421. Reference Number: [2] Reference Medline: 96146510 Reference Title: Organization of Ureaplasma urealyticum urease gene cluster Reference Title: and expression in a suppressor strain of Escherichia coli. Reference Author: Neyrolles O, Ferris S, Behbahani N, Montagnier L, Blanchard Reference Author: A; Reference Location: J Bacteriol 1996;178:647-655. Database Reference INTERPRO; IPR002639;</p>

09669980 "03060000"

Pfam	Prosite	Full Name	Description
			<p>Reference Number: [3] Reference Medline: 97325981 Reference Title: Secondary structure and tertiary fold of the human Reference Title: immunodeficiency virus protein U (Vpu) cytoplasmic domain Reference Title: in solution. Reference Author: Willbold D, Hoffmann S, Rosch P; Reference Location: Eur J Biochem 1997;245:581-588. Database Reference: SCOP; 1vpu; fa; [SCOP-USA][CATH-PDBSUM] Database Reference INTERPRO; IPR002094; Database Reference PDB; 1vpu ; 38; 81; Database reference: PFAMB; PB003303; Database reference: PFAMB; PB005882; Comment: -I- The Vpu protein contains an N-terminal transmembrane spanning region Comment: and a C-terminal cytoplasmic region. Comment: -I- The HIV-1 Vpu protein stimulates virus production by enhancing Comment: the release of viral particles from infected cells. Comment: -I- The VPU protein binds specifically to CD4. Number of members: 194</p>
XPG_N	PDOC00658	XPG protein signatures	<p>Xeroderma pigmentosum (XP) [1] is a human autosomal recessive disease, characterized by a high incidence of sunlight-induced skin cancer. People's skin cells with this condition are hypersensitive to ultraviolet light, due to defects in the incision step of DNA excision repair. There are a minimum of seven genetic complementation groups involved in this pathway: XP-A to XP-G. The defect in XP-G can be corrected by a 133 Kd nuclear protein called XPG (or XPGC) [2].</p> <p>XPG belongs to a family of proteins [2,3,4,5,6] that are composed of two main subsets:</p> <ul style="list-style-type: none"> - Subset 1, to which belongs XPG, RAD2 from budding yeast and rad13 from fission yeast. RAD2 and XPG are single-stranded DNA endonucleases [7,8]. XPG makes the 3'incision in human DNA nucleotide excision repair [9]. - Subset 2, to which belongs mouse and human FEN-1, rad2 from fission yeast, and RAD27 from budding yeast. FEN-1 is a structure-specific endonuclease. <p>In addition to the proteins listed in the above groups, this family also includes:</p> <ul style="list-style-type: none"> - Fission yeast exo1, a 5'->3' double-stranded DNA exonuclease that could act in a pathway that corrects mismatched base pairs. - Yeast EXO1 (DHS1), a protein with probably the same function as exo1. - Yeast DIN7. <p>Sequence alignment of this family of proteins reveals that similarities are largely confined to two regions. The first is located at the N-terminal extremity (N-region) and corresponds to the first 95 to 105 amino acids. The second region is internal (I-region) and found towards the C-terminus; it</p>

Pfam	Prosite	Full Name	Description
			[9] O'Donovan A., Davies A.A., Moggs J.G., West S.C., Wood R.D. Nature 371:432-435(1994):
Y_phosphatase	PDOC00323	Tyrosine specific protein phosphatases signature and profiles	<p>Tyrosine specific protein phosphatases (EC 3.1.3.48) (PTPase) [1 to 5] are enzymes that catalyze the removal of a phosphate group attached to a tyrosine residue. These enzymes are very important in the control of cell growth, proliferation, differentiation and transformation. Multiple forms of PTPase have been characterized and can be classified into two categories: soluble PTPases and transmembrane receptor proteins that contain PTPase domain(s). The currently known PTPases are listed below:</p> <p>Soluble PTPases.</p> <ul style="list-style-type: none"> - PTPN1 (PTP-1B). - PTPN2 (T-cell PTPase; TC-PTP). - PTPN3 (H1) and PTPN4 (MEG), enzymes that contain an N-terminal band 4.1-like domain (see <PDOC00566>) and could act at junctions between the membrane and cytoskeleton. - PTPN5 (STEP). - PTPN6 (PTP-1C; HCP; SHP) and PTPN11 (PTP-2C; SH-PTP3; Syp), enzymes which contain two copies of the SH2 domain at its N-terminal extremity. The <i>Drosophila</i> protein corkscrew (gene csw) also belongs to this subgroup. - PTPN7 (LC-PTP; Hematopoietic protein-tyrosine phosphatase; HePTP). - PTPN8 (70Z-PEP). - PTPN9 (MEG2). - PTPN12 (PTP-G1; PTP-P19). - Yeast PTP1. - Yeast PTP2 which may be involved in the ubiquitin-mediated protein degradation pathway. - Fission yeast pyp1 and pyp2 which play a role in inhibiting the onset of mitosis. - Fission yeast pyp3 which contributes to the dephosphorylation of cdc2. - Yeast CDC14 which may be involved in chromosome segregation. - <i>Yersinia</i> virulence plasmid PTPases (gene yopH). - <i>Autographa californica</i> nuclear polyhedrosis virus 19 Kd PTPase. <p>Dual specificity PTPases.</p> <ul style="list-style-type: none"> - DUSP1 (PTPN10; MAP kinase phosphatase-1; MKP-1); which dephosphorylates MAP kinase on both Thr-183 and Tyr-185. - DUSP2 (PAC-1), a nuclear enzyme that dephosphorylates MAP kinases ERK1 and ERK2 on both Thr and Tyr residues. - DUSP3 (VHR). - DUSP4 (HVB2). - DUSP5 (HVB3). - DUSP6 (Pyst1; MKP-3). - DUSP7 (Pyst2; MKP-X). - Yeast MSG5, a PTPase that dephosphorylates MAP kinase FUS3. - Yeast YVH1. - <i>Vaccinia</i> virus H1 PTPase; a dual specificity phosphatase. <p>Receptor PTPases.</p>

Pfam	Prosite	Full Name	Description																																																																																				
			<p>Structurally, all known receptor PTPases, are made up of a variable length extracellular domain, followed by a transmembrane region and a C-terminal catalytic cytoplasmic domain. Some of the receptor PTPases contain fibronectin type III (FN-III) repeats, immunoglobulin-like domains, MAM domains or carbonic anhydrase-like domains in their extracellular region. The cytoplasmic region generally contains two copies of the PTPase domain. The first seems to have enzymatic activity, while the second is inactive but seems to affect substrate specificity of the first. In these domains, the catalytic cysteine is generally conserved but some other, presumably important, residues are not.</p> <p>In the following table, the domain structure of known receptor PTPases is shown:</p> <table><tr><th></th><th>Extracellular</th><th colspan="4">Intracellular</th></tr><tr><th></th><th>Ig FN-3</th><th>CAH</th><th>MAM</th><th>PTPase</th><th></th></tr><tr><td>Leukocyte common antigen (LCA) (CD45)</td><td>0</td><td>2</td><td>0</td><td>0</td><td>2</td></tr><tr><td>Leukocyte antigen related (LAR)</td><td>3</td><td>8</td><td>0</td><td>0</td><td>2</td></tr><tr><td>Drosophila DLAR</td><td>3</td><td>9</td><td>0</td><td>0</td><td>2</td></tr><tr><td>Drosophila DPTP</td><td>2</td><td>2</td><td>0</td><td>0</td><td>2</td></tr><tr><td>PTP-alpha (LRP)</td><td>0</td><td>0</td><td>0</td><td>0</td><td>2</td></tr><tr><td>PTP-beta</td><td>0</td><td>16</td><td>0</td><td>0</td><td>1</td></tr><tr><td>PTP-gamma</td><td>0</td><td>1</td><td>1</td><td>0</td><td>2</td></tr><tr><td>PTP-delta</td><td>0</td><td>>7</td><td>0</td><td>0</td><td>2</td></tr><tr><td>PTP-epsilon</td><td>0</td><td>0</td><td>0</td><td>0</td><td>2</td></tr><tr><td>PTP-kappa</td><td>1</td><td>4</td><td>0</td><td>1</td><td>2</td></tr><tr><td>PTP-mu</td><td>1</td><td>4</td><td>0</td><td>1</td><td>2</td></tr><tr><td>PTP-zeta</td><td>0</td><td>1</td><td>1</td><td>0</td><td>2</td></tr></table> <p>PTPase domains consist of about 300 amino acids. There are two conserved cysteines, the second one has been shown to be absolutely required for activity. Furthermore, a number of conserved residues in its immediate vicinity have also been shown to be important.</p> <p>We derived a signature pattern for PTPase domains centered on the active site cysteine.</p> <p>There are three profiles for PTPases, the first one spans the complete domain and is not specific to any subtype. The second profile is specific to dual-specificity PTPases and the third one to the PTP subfamily.</p> <p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [LIVMF]-H-C-x(2)-G-x(3)-[STC]-[STAGPI]-x-[LIVMFY] [C is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for nine sequences. Other sequence(s) detected in SWISS-PROT 3.</p> <p>Sequences known to belong to this class detected by the 1st profile ALL. Other sequence(s) detected in SWISS-PROT 2.</p> <p>Sequences known to belong to this class detected by the 2nd</p>		Extracellular	Intracellular					Ig FN-3	CAH	MAM	PTPase		Leukocyte common antigen (LCA) (CD45)	0	2	0	0	2	Leukocyte antigen related (LAR)	3	8	0	0	2	Drosophila DLAR	3	9	0	0	2	Drosophila DPTP	2	2	0	0	2	PTP-alpha (LRP)	0	0	0	0	2	PTP-beta	0	16	0	0	1	PTP-gamma	0	1	1	0	2	PTP-delta	0	>7	0	0	2	PTP-epsilon	0	0	0	0	2	PTP-kappa	1	4	0	1	2	PTP-mu	1	4	0	1	2	PTP-zeta	0	1	1	0	2
	Extracellular	Intracellular																																																																																					
	Ig FN-3	CAH	MAM	PTPase																																																																																			
Leukocyte common antigen (LCA) (CD45)	0	2	0	0	2																																																																																		
Leukocyte antigen related (LAR)	3	8	0	0	2																																																																																		
Drosophila DLAR	3	9	0	0	2																																																																																		
Drosophila DPTP	2	2	0	0	2																																																																																		
PTP-alpha (LRP)	0	0	0	0	2																																																																																		
PTP-beta	0	16	0	0	1																																																																																		
PTP-gamma	0	1	1	0	2																																																																																		
PTP-delta	0	>7	0	0	2																																																																																		
PTP-epsilon	0	0	0	0	2																																																																																		
PTP-kappa	1	4	0	1	2																																																																																		
PTP-mu	1	4	0	1	2																																																																																		
PTP-zeta	0	1	1	0	2																																																																																		

Pfam	Prosite	Full Name	Description
			<p>profile ALL dual type PTPases. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Sequences known to belong to this class detected by the 3rd profile ALL PTP type PTPases. Other sequence(s) detected in SWISS-PROT NONE.</p> <p>Note the M-phase inducer phosphatases (cdc25-type phosphatase) are tyrosine- protein phosphatases that are not structurally related to the above PTPases.</p> <p>Note this documentation entry is linked to both a signature pattern and to profiles. As profiles are much more sensitive than the pattern, you should use them if you have access to the necessary software tools to do so. Last update July 1999 / Text revised. References [1] Fischer E.H., Charbonneau H., Tonks N.K. Science 253:401-406(1991).</p> <p>[2] Charbonneau H., Tonks N.K. Annu. Rev. Cell Biol. 8:463-493(1992).</p> <p>[3] Trowbridge I.S. J. Biol. Chem. 266:23517-23520(1991).</p> <p>[4] Tonks N.K., Charbonneau H. Trends Biochem. Sci. 14:497-500(1989).</p> <p>[5] Hunter T. Cell 58:1013-1016(1989).</p>
Zein		Zein seed storage protein	<p>Accession number: PF01559 Definition: Zein seed storage protein Author: Bateman A Alignment method of seed: Clustalw Source of seed members: Pfam-B_181 (release 4.0) Gathering cutoffs: -21 -21 Trusted cutoffs: 4.60 4.60 Noise cutoffs: -46.60 -46.60 HMM build command line: hmmbuild -F HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 93197294 Reference Title: Studies of the zein-like alpha-prolamins based on an Reference Title: analysis of amino acid sequences: implications for their Reference Title: evolution and three-dimensional structure. Reference Author: Garratt R, Oliva G, Caracelli I, Leite A, Arruda P; Reference Location: Proteins 1993;15:88-99. Database Reference: INTERPRO; IPR002530; Comment: Zeins are seed storage proteins. They are unusually rich in Comment: glutamine, proline, alanine, and leucine residues and their Comment: sequences show a series of tandem repeats [1]. Number of members: 48</p>
zf-AN1		AN1-like Zinc finger	<p>Accession number: PF01428 Definition: AN1-like Zinc finger Author: Bateman A, SMART Alignment method of seed: Manual Source of seed members: SMART Gathering cutoffs: 16 16 Trusted cutoffs: 16.40 16.40</p>

	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	2101	2102	2103	2104	2105	2106	2107	2108	2109	2110	2111	2112	2113	2114	2115	2116	2117	2118	2119	2120	2121	2122	2123	2124	2125	2126	2127	2128	2129	2130	2131	2132	2133	2134	2135	2136	2137	2138	2139	2140	2141	2142	2143	2144	2145	2146	2147	2148	2149	2150	2151	2152	2153	2154	2155	2156	2157	2158	2159	2160	2161	2162	2163	2164	2165	2166	2167	2168	2169	2170	2171	2172	2173	2174	2175	2176	2177	2178	2179	2180	2181	2182	2183	2184	2185	2186	2187	2188	2189	2190	2191	2192	2193	2194	2195	2196	2197	2198	2199	2200	2201	2202	2203	2204	2205	2206	2207	2208	2209	2210	2211	2212	2213	2214	2215	2216	2217	2218	2219	2220	2221	2222	2223	2224	2225	2226	2227	2228	2229	2230	2231	2232	2233	2234	2235	2236	2237	2238	2239	2240	2241	2242	2243	2244	2245	2246	2247	2248	2249	2250	2251	2252	2253	2254	2255	2256	2257	2258	2259	2260	2261	2262	2263	2264	2265	2266	2267	2268	2269	2270	2271	2272	2273	2274	2275	2276	2277	2278	2279	2280	2281	2282	2283	2284	2285	2286	2287	2288	2289	2290	2291	2292	2293	2294	2295	2296	2297	2298	2299	2300	2301	2302	2303	2304	2305	2306	2307	2308	2309	2310	2311	2312	2313	2314	2315	2316	2317	2318	2319	2320	2321	2322	2323	2324	2325	2326	2327	2328	2329	2330	2331	2332	2333	2334	2335	2336	2337	2338	2339	2340	2341	2342	2343	2344	2345	2346	2347	2348	2349	2350	2351	2352	2353	2354	2355	2356	2357	2358	2359	2360	2361	2362	2363	2364	2365	2366	2367	2368	2369	2370	2371	2372	2373	2374	2375	2376	2377	2378	2379	2380	2381	2382	2383	2384	2385	2386	2387	2388	2389	2390	2391	2392	2393	2394	2395	2396	2397	2398	2399	2400	2401	2402	2403	2404	2405	2406	2407	2408	2409	2410	2411	2412	2413	2414	2415	2416	2417	2418	2419	2420	2421	2422	2
--	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	---

Pfam	Prosite	Full Name	Description
			<p>Noise cutoffs: 7.30 7.30</p> <p>HMM build command line: hmmbuild HMM SEED</p> <p>HMM build command line: hmmcalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 93292985</p> <p>Reference Title: Two related localized mRNAs from <i>Xenopus laevis</i> encode</p> <p>Reference Title: ubiquitin-like fusion proteins.</p> <p>Reference Author: Linnen JM, Bailey CP, Weeks DL;</p> <p>Reference Location: Gene 1993;128:181-188.</p> <p>Database reference: SMART; ZnF_AN1;</p> <p>Database Reference: INTERPRO; IPR000058;</p> <p>Comment: Zinc finger at the C-terminus of An1</p> <p>Swiss:Q91889, a ubiquitin-like</p> <p>Comment: protein in <i>Xenopus laevis</i>.</p> <p>Comment: The following pattern describes the zinc finger.</p> <p>Comment: C-X2-C-X(9-12)-C-X(1-2)-C-X4-C-X2-H-X5-H-X-C</p> <p>Comment: Where X can be any amino acid, and numbers in brackets</p> <p>Comment: indicate the number of residues.</p> <p>Number of members: 18</p>
zf-B_box	PDOC50015	B-box zinc finger	<p>Accession number: PF00643</p> <p>Definition: B-box zinc finger.</p> <p>Author: Bateman A</p> <p>Alignment method of seed: pftools</p> <p>Source of seed members: Prosite</p> <p>Gathering cutoffs: 25 25</p> <p>Trusted cutoffs: 26.00 26.00</p> <p>Noise cutoffs: 24.50 29.90</p> <p>HMM build command line: hmmbuild HMM SEED</p> <p>HMM build command line: hmmcalibrate --seed 0 HMM</p> <p>Database Reference: SCOP; 1fre; fa; [SCOP-USA][CATH-PDBSUM]</p> <p>Database reference: PROSITE_PROFILE; PS50119;</p> <p>Database Reference: PROSITE; PDOC50015</p> <p>Database Reference: INTERPRO; IPR002991;</p> <p>Database Reference: PDB; 1fre ; 4; 42;</p> <p>Database reference: PFAMB; PB002777;</p> <p>Database reference: PFAMB; PB010625;</p> <p>Database reference: PFAMB; PB041771;</p> <p>Number of members: 44</p>
zf-CONSTANS		CONSTANS family zinc finger	<p>Accession number: PF01760</p> <p>Definition: CONSTANS family zinc finger</p> <p>Author: Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_1072 (release 4.2)</p> <p>Gathering cutoffs: 25 10</p> <p>Trusted cutoffs: 76.10 17.20</p> <p>Noise cutoffs: 9.70 9.70</p> <p>HMM build command line: hmmbuild HMM SEED</p> <p>HMM build command line: hmmcalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 95211836</p> <p>Reference Title: The CONSTANS gene of <i>Arabidopsis</i> promotes flowering and</p> <p>Reference Title: encodes a protein showing similarities to zinc finger</p> <p>Reference Title: transcription factors.</p> <p>Reference Author: Putterill J, Robson F, Lee K, Simon R, Coupland G;</p> <p>Reference Location: Cell 1995;80:847-857.</p> <p>Database Reference: INTERPRO; IPR002926;</p> <p>Number of members: 45</p>
zf-DHHC		DHHC zinc finger domain	<p>Accession number: PF01529</p> <p>Definition: DHHC zinc finger domain</p> <p>Author: Bateman A</p> <p>Alignment method of seed: Clustalw</p> <p>Source of seed members: Pfam-B_945 (release 4.0)</p> <p>Gathering cutoffs: 22 22</p>

Pfam	Prosite	Full Name	Description
			<p>Trusted cutoffs: 22.40 22.40 Noise cutoffs: -22.40 -22.40 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 99250263 Reference Title: The drosophila STAM gene homolog is in a tight gene Reference Title: cluster, and its expression correlates to that of the Reference Title: adjacent gene ial. Reference Author: Mesilaty-Gross S, Reich A, Motro B, Wides R; Reference Location: Gene 1999;231:173-186. Reference Number: [2] Reference Medline: 97315340 Reference Title: Variations of the C2H2 zinc finger motif in the yeast Reference Title: genome and classification of yeast zinc finger proteins. Reference Author: Bohm S, Frishman D, Mewes HW; Reference Location: Nucleic Acids Res 1997;25:2464-2469. Reference Number: [3] Reference Medline: 99321009 Reference Title: The DHHC domain: a new highly conserved cysteine-rich Reference Title: motif. Reference Author: Putilina T, Wong P, Gentleman S; Reference Location: Mol Cell Biochem 1999;195:219-226. Reference Number: [4] Reference Medline: 10490616 Reference Title: Erf2, a Novel Gene Product That Affects the Localization Reference Title: and Palmitoylation of Ras2 in Saccharomyces cerevisiae. Reference Author: Bartels DJ, Mitchell DA, Dong X, Deschenes RJ; Reference Location: Mol Cell Biol 1999;19:6775-6787. Database Reference: INTERPRO; IPR001594; Comment: This domain is also known as NEW1 [2]. This domain is Comment: predicted to be a zinc binding domain. The function Comment: of this domain is unknown, but it has been predicted to Comment: be involved in protein-protein or protein-DNA interactions [3]. Number of members: 34</p>
zf-MYND		MYND finger	<p>Accession number: PF01753 Definition: MYND finger Author: Bateman A Alignment method of seed: Manual Source of seed members: Bateman A Gathering cutoffs: 11 11 Trusted cutoffs: 17.30 17.30 Noise cutoffs: 5.50 5.50 HMM build command line: hmmbuild HMM SEED HMM build command line: hmmscalibrate --seed 0 HMM Reference Number: [1] Reference Medline: 96203118 Reference Title: DEAF-1, a novel protein that binds an essential region in a Reference Title: Deformed response element. Reference Author: Gross CT, McGinnis W; Reference Location: EMBO J 1996;15:1961-1970. Reference Number: [2] Reference Medline: 98079069 Reference Title: Molecular cloning, sequence analysis, expression, and Reference Title: tissue distribution of suppressin, a novel suppressor of Reference Title: cell cycle entry. Reference Author: LeBoeuf RD, Ban EM, Green MM, Stone</p>

Pfam	Prosite	Full Name	Description
			<p>Description of pattern(s) and/or profile(s)</p> <p>Consensus pattern [PK]-x-[LIVMFY]-x-[LIVMFY]-x(4)-H-[STAG]-x-E-x-[LIVM]-[STAG]-x(6)-[LIVMFYTA] [H and E are zinc ligands] Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT <i>Bacillus sphaericus</i> endopeptidase I which hydrolyses the gamma-D-Glu-(L)meso-diaminopimelic acid bond of spore cortex peptidoglycan [6] and which is possibly distantly related to zinc carboxypeptidases.</p> <p>Consensus pattern H-[STAG]-x(3)-[LIVME]-x(2)-[LIVMFYW]-P-[FYW] [H is a zinc ligand] Sequences known to belong to this class detected by the pattern ALL.</p> <p>Other sequence(s) detected in SWISS-PROT 40.</p> <p>Note if a protein includes both signatures, the probability of it being a eukaryotic zinc carboxypeptidase is 100%</p> <p>Note these proteins belong to families M14A/M14B in the classification of peptidases [7,E1].</p> <p>Last update November 1995 / Patterns and text revised.</p> <p>References</p> <p>[1] Tan F., Chan S.J., Steiner D.F., Schilling J.W., Skidgel R.A. J. Biol. Chem. 264:13165-13170(1989).</p> <p>[2] Reynolds D.S., Stevens R.L., Gurley D.S., Lane W.S., Austen K.F., Serafin W.E. J. Biol. Chem. 264:20094-20099(1989).</p> <p>[3] Narahashi Y. J. Biochem. 107:879-886(1990).</p> <p>[4] Teplyakov A., Polyakov K., Obmolova G., Strokopytov B., Kuranova I., Osterman A.L., Grishin N.V., Smulevitch S.V., Zagnitko O.P., Galperina O.V., Matz M.V., Stepanov V.M. Eur. J. Biochem. 208:281-288(1992).</p> <p>[5] He G.-P., Muise A., Li A.W., Ro H.-S. Nature 378:92-96(1995).</p> <p>[6] Hourdou M.-L., Guinand M., Vacheron M.J., Michel G., Denoroy L., Duez C.M., Englebert S., Joris B., Weber G., Ghuysen J.-M. Biochem. J. 292:563-570(1993).</p> <p>[7] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-228(1995).</p> <p>[E1] http://www.expasy.ch/cgi-bin/lists?peptidas.txt</p>
ZZ		Zinc finger present in dystrophin, CBP/p300	<p>Accession number: PF00569</p> <p>Definition: Zinc finger present in dystrophin, CBP/p300</p> <p>Author: SMART</p> <p>Alignment method of seed: Manual</p> <p>Source of seed members: Alignment kindly provided by SMART</p> <p>Gathering cutoffs: 14 14</p> <p>Trusted cutoffs: 14.60 14.60</p> <p>Noise cutoffs: 10.90 10.90</p> <p>HMM build command line: hmmbuild HMM SEED</p> <p>HMM build command line: hmmscalibrate --seed 0 HMM</p> <p>Reference Number: [1]</p> <p>Reference Medline: 96402609</p> <p>Reference Title: ZZ and TAZ: new putative zinc fingers in</p>

00639980-101300

Pfam	Prosite	Full Name	Description
			dystrophin and Reference Title: other proteins. Reference Author: Ponting CP, Blake DJ, Davies KE, Kendrick-Jones J, Winder Reference Author: SJ; Reference Location: Trends Biochem Sci 1996;21:11-13. Database Reference: EXPERT; Chris.Ponting@human- anatomy.oxford.ac.uk; Database Reference INTERPRO; IPR000433; Database reference: PFAMB; PB041629; Comment: ZZ in dystrophin binds calmodulin Comment: Putative zinc finger; binding not yet shown. Number of members: 87

09689980-101300

AA. Activities of Polypeptides Comprising Signal Peptides

Polypeptides comprising signal peptides are a family of proteins that are typically
5 targeted to (1) a particular organelle or intracellular compartment, (2) interact with a
particular molecule or (3) for secretion outside of a host cell. Example of polypeptides
comprising signal peptides include, without limitation, secreted proteins, soluble proteins,
receptors, proteins retained in the ER, etc.

10 These proteins comprising signal peptides are useful to modulate ligand-receptor
interactions, cell-to-cell communication, signal transduction, intracellular communication,
and activities and/or chemical cascades that take part in an organism outside or within of any
particular cell.

15 One class of such proteins are soluble proteins which are transported out of the cell.
These proteins can act as ligands that bind to receptor to trigger signal transduction or to
permit communication between cells.

20 Another class is receptor proteins which also comprise a retention domain that lodges
the receptor protein in the membrane when the cell transports the receptor to the surface of
the cell. Like the soluble ligands, receptors can also modulate signal transduction and
communication between cells.

25 In addition the signal peptide itself can serve as a ligand for some receptors. An
example is the interaction of the ER targeting signal peptide with the signal recognition
particle (SRP). Here, the SRP binds to the signal peptide, halting translation, and the
resulting SRP complex then binds to docking proteins located on the surface of the ER,
prompting transfer of the protein into the ER.

30 A description of signal peptide residue composition is described below in Subsection
IV.C.1.

09689980-1047-1078

III. Methods of Modulating Polypeptide Production

It is contemplated that polynucleotides of the invention can be incorporated into a host cell or in-vitro system to modulate polypeptide production. For instance, the SDFs prepared as described herein can be used to prepare expression cassettes useful in a number of techniques for suppressing or enhancing expression.

An example are polynucleotides comprising sequences to be transcribed, such as coding sequences, of the present invention can be inserted into nucleic acid constructs to modulate polypeptide production. Typically, such sequences to be transcribed are heterologous to at least one element of the nucleic acid construct to generate a chimeric gene or construct.

Another example of useful polynucleotides are nucleic acid molecules comprising regulatory sequences of the present invention. Chimeric genes or constructs can be generated when the regulatory sequences of the invention linked to heterologous sequences in a vector construct. Within the scope of invention are such chimeric gene and/or constructs.

Also within the scope of the invention are nucleic acid molecules, whereof at least a part or fragment of these DNA molecules are presented in Tables 1 and 2 of the present application, and wherein the coding sequence is under the control of its own promoter and/or its own regulatory elements. Such molecules are useful for transforming the genome of a host cell or an organism regenerated from said host cell for modulating polypeptide production.

Additionally, a vector capable of producing the oligonucleotide can be inserted into the host cell to deliver the oligonucleotide.

More detailed description of components to be included in vector constructs are described both above and below.

Whether the chimeric vectors or native nucleic acids are utilized, such polynucleotides can be incorporated into a host cell to modulate polypeptide production. Native genes and/or nucleic acid molecules can be effective when exogenous to the host cell.

Methods of modulating polypeptide expression includes, without limitation:

Suppression methods, such as

Antisense

Ribozymes

Co-suppression

Insertion of Sequences into the Gene to be Modulated

Regulatory Sequence Modulation.

as well as Methods for Enhancing Production, such as
Insertion of Exogenous Sequences; and
Regulatory Sequence Modulation.

5 III.A. Suppression

Expression cassettes of the invention can be used to suppress expression of
endogenous genes which comprise the SDF sequence. Inhibiting expression can be useful,
for instance, to tailor the ripening characteristics of a fruit (Oeller et al., *Science* 254:437
(1991)) or to influence seed size_(WO98/07842) or to provoke cell ablation (Mariani et al.,
10 Nature 357: 384-387 (1992)).

As described above, a number of methods can be used to inhibit gene expression in
plants, such as antisense, ribozyme, introduction of exogenous genes into a host cell,
insertion of a polynucleotide sequence into the coding sequence and/or the promoter of the
endogenous gene of interest, and the like.

15 III.A.1. Antisense

An expression cassette as described above can be transformed into host cell or
plant to produce an antisense strand of RNA. For plant cells, antisense RNA inhibits gene
expression by preventing the accumulation of mRNA which encodes the enzyme of interest, *see*,
e.g., Sheehy et al., *Proc. Nat. Acad. Sci. USA*, 85:8805 (1988), and Hiatt et al., U.S. Patent No.
20 4,801,340.

 III.A.2. Ribozymes

Similarly, ribozyme constructs can be transformed into a plant to cleave mRNA
and down-regulate translation.

 III.A.3. Co-Suppression

25 Another method of suppression is by introducing an exogenous copy of the gene
to be suppressed. Introduction of expression cassettes in which a nucleic acid is configured in
the sense orientation with respect to the promoter has been shown to prevent the accumulation of
mRNA. A detailed description of this method is described above.

 III.A.4. Insertion of Sequences into the Gene to be Modulated

05689980-101300

Yet another means of suppressing gene expression is to insert a polynucleotide into the gene of interest to disrupt transcription or translation of the gene.

Homologous recombination could be used to target a polynucleotide insert to a gene using the Cre-Lox system (A.C. Vergunst et al., *Nucleic Acids Res.* 26:2729 (1998), A.C. Vergunst et al., *Plant Mol. Biol.* 38:393 (1998), H. Albert et al., *Plant J.* 7:649 (1995)).

In addition, random insertion of polynucleotides into a host cell genome can also be used to disrupt the gene of interest. Azpiroz-Leehan et al., *Trends in Genetics* 13:152 (1997). In this method, screening for clones from a library containing random insertions is preferred for identifying those that have polynucleotides inserted into the gene of interest. Such screening can be performed using probes and/or primers described above based on sequences from Tables 1 and 2, fragments thereof, and substantially similar sequence thereto. The screening can also be performed by selecting clones or any transgenic plants having a desired phenotype.

III.A.5. Regulatory Sequence Modulation

The SDFs described in Tables 1 and 2, and fragments thereof are examples of nucleotides of the invention that contain regulatory sequences that can be used to suppress or inactivate transcription and/or translation from a gene of interest as discussed in I.C.5.

III.A.6. Genes Comprising Dominant-Negative Mutations

When suppression of production of the endogenous, native protein is desired it is often helpful to express a gene comprising a dominant negative mutation. Production of protein variants produced from genes comprising dominant negative mutations is a useful tool for research. Genes comprising dominant negative mutations can produce a variant polypeptide which is capable of competing with the native polypeptide, but which does not produce the native result. Consequently, over expression of genes comprising these mutations can titrate out an undesired activity of the native protein. For example, The product from a gene comprising a dominant negative mutation of a receptor can be used to constitutively activate or suppress a signal transduction cascade, allowing examination of the phenotype and thus the trait(s) controlled by that receptor and pathway. Alternatively, the protein arising from the gene comprising a dominant-negative mutation can be an inactive enzyme still capable of binding to the same substrate as the native protein and therefore competes with such native protein.

Products from genes comprising dominant-negative mutations can also act upon the native protein itself to prevent activity. For example, the native protein may be active only as a homo-multimer or as one subunit of a hetero-multimer. Incorporation of an inactive subunit into the multimer with native subunit(s) can inhibit activity.

Thus, gene function can be modulated in host cells of interest by insertion into these cells vector constructs comprising a gene comprising a dominant-negative mutation.

III.B. Enhanced Expression

Enhanced expression of a gene of interest in a host cell can be accomplished by either (1) insertion of an exogenous gene; or (2) promoter modulation.

III.B.1. Insertion of an Exogenous Gene

Insertion of an expression construct encoding an exogenous gene can boost the number of gene copies expressed in a host cell.

Such expression constructs can comprise genes that either encode the native protein that is of interest or that encode a variant that exhibits enhanced activity as compared to the native protein. Such genes encoding proteins of interest can be constructed from the sequences from Tables 1 and 2, fragments thereof, and substantially similar sequence thereto.

Such an exogenous gene can include either a constitutive promoter permitting expression in any cell in a host organism or a promoter that directs transcription only in particular cells or times during a host cell life cycle or in response to environmental stimuli.

III.B.2. Regulatory Sequence Modulation

The SDFs of Tables 1 and 2, and fragments thereof, contain regulatory sequences that can be used to enhance expression of a gene of interest. For example, some of these sequences contain useful enhancer elements. In some cases, duplication of enhancer elements or insertion of exogenous enhancer elements will increase expression of a desired gene from a particular promoter. As other examples, all II promoters require binding of a regulatory protein to be activated, while some promoters may need a protein that signals a promoter binding protein to expose a polymerase binding site. In either case, over-production of such proteins can be used to enhance expression of a gene of interest by increasing the activation time of the promoter.

Such regulatory proteins are encoded by some of the sequences in Tables 1 and 2, fragments thereof, and substantially similar sequences thereto.

Coding sequences for these proteins can be constructed as described above.

IV. Gene Constructs and Vector Construction

To use isolated SDFs of the present invention or a combination of them or parts and/or mutants and/or fusions of said SDFs in the above techniques, recombinant DNA vectors which comprise said SDFs and are suitable for transformation of cells, such as plant cells, are usually prepared. The SDF construct can be made using standard recombinant DNA techniques (Sambrook et al. 1989) and can be introduced to the species of interest by *Agrobacterium*-mediated transformation or by other means of transformation (e.g., particle gun bombardment) as referenced below.

The vector backbone can be any of those typical in the art such as plasmids, viruses, artificial chromosomes, BACs, YACs and PACs and vectors of the sort described by

- (a) **BAC:** Shizuya et al., Proc. Natl. Acad. Sci. USA 89: 8794-8797 (1992); Hamilton et al., Proc. Natl. Acad. Sci. USA 93: 9975-9979 (1996);
- (b) **YAC:** Burke et al., Science 236:806-812 (1987);.
- (c) **PAC:** Sternberg N. et al., Proc Natl Acad Sci U S A. Jan;87(1):103-7 (1990);
- (d) **Bacteria-Yeast Shuttle Vectors:** Bradshaw et al., Nucl Acids Res 23: 4850-4856 (1995);
- (e) **Lambda Phage Vectors:** Replacement Vector, e.g., Frischauf et al., J. Mol Biol 170: 827-842 (1983); or Insertion vector, e.g., Huynh et al., In: Glover NM (ed) DNA Cloning: A practical Approach, Vol.1 Oxford: IRL Press (1985);
- (f) **T-DNA gene fusion vectors :**Walden et al., Mol Cell Biol 1: 175-194 (1990); and
- (g) **Plasmid vectors:** Sambrook et al., infra.

Typically, a vector will comprise the exogenous gene, which in its turn comprises an SDF of the present invention to be introduced into the genome of a host cell, and which gene may be an antisense construct, a ribozyme construct chimera, or a coding sequence with any desired transcriptional and/or translational regulatory sequences, such as promoters, UTRs, and 3' end termination sequences. Vectors of the invention can also include origins of replication, scaffold attachment regions (SARs), markers, homologous sequences, introns, etc.

A DNA sequence coding for the desired polypeptide, for example a cDNA sequence encoding a full length protein, will preferably be combined with transcriptional and translational

initiation regulatory sequences which will direct the transcription of the sequence from the gene in the intended tissues of the transformed plant.

For example, for over-expression, a plant promoter fragment may be employed that will direct transcription of the gene in all tissues of a regenerated plant. Alternatively, the plant promoter may direct transcription of an SDF of the invention in a specific tissue (tissue-specific promoters) or may be otherwise under more precise environmental control (inducible promoters).

If proper polypeptide production is desired, a polyadenylation region at the 3'-end of the coding region is typically included. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA.

The vector comprising the sequences from genes or SDF or the invention may comprise a marker gene that confers a selectable phenotype on plant cells. The vector can include promoter and coding sequence, for instance. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosulfuron or phosphinotricin.

IV.A. Coding Sequences

Generally, the sequence in the transformation vector and to be introduced into the genome of the host cell does not need to be absolutely identical to an SDF of the present invention. Also, it is not necessary for it to be full length, relative to either the primary transcription product or fully processed mRNA. Furthermore, the introduced sequence need not have the same intron or exon pattern as a native gene. Also, heterologous non-coding segments can be incorporated into the coding sequence without changing the desired amino acid sequence of the polypeptide to be produced.

IV.B. Promoters

As explained above, introducing an exogenous SDF from the same species or an orthologous SDF from another species can modulate the expression of a native gene corresponding to that SDF of interest. Such an SDF construct can be under the control of either a constitutive promoter or a highly regulated inducible promoter (*e.g.*, a copper inducible promoter). The promoter of interest can initially be either endogenous or heterologous to the species in question. When re-introduced into the genome of said species, such promoter becomes exogenous to said species. Over-expression of an SDF transgene can

lead to co-suppression of the homologous endogenous sequence thereby creating some alterations in the phenotypes of the transformed species as demonstrated by similar analysis of the chalcone synthase gene (Napoli et al., *Plant Cell* 2:279 (1990) and van der Krol et al., *Plant Cell* 2:291 (1990)). If an SDF is found to encode a protein with desirable characteristics, its over-production can be controlled so that its accumulation can be manipulated in an organ- or tissue-specific manner utilizing a promoter having such specificity.

Likewise, if the promoter of an SDF (or an SDF that includes a promoter) is found to be tissue-specific or developmentally regulated, such a promoter can be utilized to drive or facilitate the transcription of a specific gene of interest (e.g., seed storage protein or root-specific protein). Thus, the level of accumulation of a particular protein can be manipulated or its spatial localization in an organ- or tissue- specific manner can be altered.

IV. C Signal Peptides

SDFs of the present invention containing signal peptides are indicated in Tables 1 and 2. In some cases it may be desirable for the protein encoded by an introduced exogenous or orthologous SDF to be targeted (1) to a particular organelle intracellular compartment, (2) to interact with a particular molecule such as a membrane molecule or (3) for secretion outside of the cell harboring the introduced SDF. This will be accomplished using a signal peptide.

Signal peptides direct protein targeting, are involved in ligand-receptor interactions and act in cell to cell communication. Many proteins, especially soluble proteins, contain a signal peptide that targets the protein to one of several different intracellular compartments. In plants, these compartments include, but are not limited to, the endoplasmic reticulum (ER), mitochondria, plastids (such as chloroplasts), the vacuole, the Golgi apparatus, protein storage vesicles (PSV) and, in general, membranes. Some signal peptide sequences are conserved, such as the Asn-Pro-Ile-Arg amino acid motif found in the N-terminal propeptide signal that targets proteins to the vacuole (Marty (1999) *The Plant Cell* 11: 587-599). Other signal peptides do not have a consensus sequence *per se*, but are largely composed of hydrophobic amino acids, such as those signal peptides targeting proteins to the ER (Vitale and Denecke (1999) *The Plant Cell* 11: 615-628). Still others do not appear to contain either a consensus sequence or an identified common secondary sequence, for instance the chloroplast stromal targeting signal peptides (Keegstra and Cline (1999) *The Plant Cell* 11: 557-570). Furthermore, some targeting peptides are bipartite, directing proteins first to an organelle and then to a membrane within the organelle (e.g. within the thylakoid lumen of the

chloroplast; see Keegstra and Cline (1999) *The Plant Cell* 11: 557-570). In addition to the diversity in sequence and secondary structure, placement of the signal peptide is also varied. Proteins destined for the vacuole, for example, have targeting signal peptides found at the N-terminus, at the C-terminus and at a surface location in mature, folded proteins. Signal peptides also serve as ligands for some receptors.

These characteristics of signal proteins can be used to more tightly control the phenotypic expression of introduced SDFs. In particular, associating the appropriate signal sequence with a specific SDF can allow sequestering of the protein in specific organelles (plastids, as an example), secretion outside of the cell, targeting interaction with particular receptors, etc. Hence, the inclusion of signal proteins in constructs involving the SDFs of the invention increases the range of manipulation of SDF phenotypic expression. The nucleotide sequence of the signal peptide can be isolated from characterized genes using common molecular biological techniques or can be synthesized in vitro.

In addition, the native signal peptide sequences, both amino acid and nucleotide, described in Tables 1 and 2 can be used to modulate polypeptide transport. Further variants of the native signal peptides described in Tables 1 and 2 are contemplated. Insertions, deletions, or substitutions can be made. Such variants will retain at least one of the functions of the native signal peptide as well as exhibiting some degree of sequence identity to the native sequence.

Also, fragments of the signal peptides of the invention are useful and can be fused with other signal peptides of interest to modulate transport of a polypeptide.

V. Transformation Techniques

A wide range of techniques for inserting exogenous polynucleotides are known for a number of host cells, including, without limitation, bacterial, yeast, mammalian, insect and plant cells.

Techniques for transforming a wide variety of higher plant species are well known and described in the technical and scientific literature. See, e.g. Weising et al., *Ann. Rev. Genet.* 22:421 (1988); and Christou, *Euphytica*, v. 85, n.1-3:13-27, (1995).

DNA constructs of the invention may be introduced into the genome of the desired plant host by a variety of conventional techniques. For example, the DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment. Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and

introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria (McCormac et al., *Mol. Biotechnol.* 8:199 (1997); Hamilton, *Gene* 200:107 (1997)); Salomon et al. *EMBO J.* 3:141 (1984); Herrera-Estrella et al. *EMBO J.* 2:987 (1983).

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski et al. *EMBO J.* 3:2717 (1984). Electroporation techniques are described in Fromm et al. *Proc. Natl Acad. Sci. USA* 82:5824 (1985). Ballistic transformation techniques are described in Klein et al. *Nature* 327:773 (1987). *Agrobacterium tumefaciens*-mediated transformation techniques, including disarming and use of binary or co-integrate vectors, are well described in the scientific literature. See, for example Hamilton, CM., *Gene* 200:107 (1997); Müller et al. *Mol. Gen. Genet.* 207:171 (1987); Komari et al. *Plant J.* 10:165 (1996); Venkateswarlu et al. *Biotechnology* 2:1103 (1991) and Gleave, AP., *Plant Mol. Biol.* 20:1203 (1992); Graves and Goldman, *Plant Mol. Biol.* 7:34 (1986) and Gould et al., *Plant Physiology* 95:426 (1991).

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant that possesses the transformed genotype and thus the desired phenotype such as seedlessness. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans et al., *Protoplasts Isolation and Culture* in "Handbook of Plant Cell Culture," pp. 124-176, MacMillan Publishing Company, New York, 1983; and Binding, *Regeneration of Plants, Plant Protoplasts*, pp. 21-73, CRC Press, Boca Raton, 1988. Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee et al. *Ann. Rev. of Plant Phys.* 38:467 (1987). Regeneration of monocots (rice) is described by Hosoyama et al. (*Biosci. Biotechnol. Biochem.* 58:1500 (1994)) and by Ghosh et al. (*J. Biotechnol.* 32:1 (1994)). The nucleic acids of the invention can be used to confer desired traits on essentially any plant.

Thus, the invention has use over a broad range of plants, including species from the genera *Anacardium*, *Arachis*, *Asparagus*, *Atropa*, *Avena*, *Brassica*, *Citrus*, *Citrullus*, *Capsicum*, *Carthamus*, *Cocos*, *Coffea*, *Cucumis*, *Cucurbita*, *Daucus*, *Elaeis*, *Fragaria*, *Glycine*, *Gossypium*, *Helianthus*, *Heterocallis*, *Hordeum*, *Hyoscyamus*, *Lactuca*, *Linum*, *Lolium*, *Lupinus*,

Lycopersicon, Malus, Manihot, Majorana, Medicago, Nicotiana, Olea, Oryza, Panieum, Pannesetum, Persea, Phaseolus, Pistachia, Pisum, Pyrus, Prunus, Raphanus, Ricinus, Secale, Senecio, Sinapis, Solanum, Sorghum, Theobromus, Trigonella, Triticum, Vicia, Vitis, Vigna, and, Zea.

5 One of skill will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

10 The particular sequences of SDFs identified are provided in the attached Tables 1 and 2. One of ordinary skill in the art, having this data, can obtain cloned DNA fragments, synthetic DNA fragments or polypeptides constituting desired sequences by recombinant methodology known in the art or described herein.

EXAMPLES

15 The invention is illustrated by way of the following examples. The invention is not limited by these examples as the scope of the invention is defined solely by the claims following.

EXAMPLE 1: cDNA PREPARATION

20 A number of the nucleotide sequences disclosed in Tables 1 and 2 herein as representative of the SDFs of the invention can be obtained by sequencing genomic DNA (gDNA) and/or cDNA from corn plants grown from HYBRID SEED # 35A19, purchased from Pioneer Hi-Bred International, Inc., Supply Management, P.O. Box 256, Johnston, Iowa 50131-0256.

25 A number of the nucleotide sequences disclosed in Tables 1 and 2 herein as representative of the SDFs of the invention can also be obtained by sequencing genomic DNA from *Arabidopsis thaliana*, Wassilewskija ecotype or by sequencing cDNA obtained from mRNA from such plants as described below. This is a true breeding strain. Seeds of the plant are available from the Arabidopsis Biological Resource Center at the Ohio State University, under the accession number CS2360. Seeds of this plant were deposited under the terms and conditions of the Budapest Treaty at the American Type Culture Collection, 30 Manassas, VA on August 31, 1999, and were assigned ATCC No. PTA-595.

Other methods for cloning full-length cDNA are described, for example, by Seki et al., *Plant Journal* 15:707-720 (1998) High-efficiency cloning of Arabidopsis full-length

09689980-101300

cDNA by biotinylated Cap trapper"; Maruyama et al., *Gene* 138:171 (1994) Oligo-capping a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides"; and WO 96/34981.

5 Tissues were, or each organ was, individually pulverized and frozen in liquid nitrogen. Next, the samples were homogenized in the presence of detergents and then centrifuged. The debris and nuclei were removed from the sample and more detergents were added to the sample. The sample was centrifuged and the debris was removed. Then the sample was applied to a 2M sucrose cushion to isolate polysomes. The RNA was isolated by treatment with detergents and proteinase K followed by ethanol precipitation and
10 centrifugation. The polysomal RNA from the different tissues was pooled according to the following mass ratios: 15/15/1 for male inflorescences, female inflorescences and root, respectively. The pooled material was then used for cDNA synthesis by the methods described below.

15 Starting material for cDNA synthesis for the exemplary corn cDNA clones with sequences presented in Tables 1 and 2 was poly(A)-containing polysomal mRNAs from inflorescences and root tissues of corn plants grown from HYBRID SEED # 35A19. Male inflorescences and female (pre-and post-fertilization) inflorescences were isolated at various stages of development. Selection for poly(A) containing polysomal RNA was done using oligo d(T) cellulose columns, as described by Cox and Goldberg, *Plant Molecular Biology: A Practical Approach*", pp. 1-35, Shaw ed., c. 1988 by IRL, Oxford. The quality and the
20 integrity of the polyA+ RNAs were evaluated.

25 Starting material for cDNA synthesis for the exemplary *Arabidopsis* cDNA clones with sequences presented in Tables 1 and 2 was polysomal RNA isolated from the top-most inflorescence tissues of *Arabidopsis thaliana* Wassilewskija (Ws.) and from roots of *Arabidopsis thaliana* Landsberg erecta (L. er.), also obtained from the Arabidopsis Biological Resource Center. Nine parts inflorescence to every part root was used, as measured by wet mass. Tissue was pulverized and exposed to liquid nitrogen. Next, the sample was homogenized in the presence of detergents and then centrifuged. The debris and
30 nuclei were removed from the sample and more detergents were added to the sample. The sample was centrifuged and the debris was removed and the sample was applied to a 2M sucrose cushion to isolate polysomal RNA. Cox et al., *Plant Molecular Biology: A Practical Approach*", pp. 1-35, Shaw ed., c. 1988 by IRL, Oxford. The polysomal RNA was used

for cDNA synthesis by the methods described below. Polysomal mRNA was then isolated as described above for corn cDNA. The quality of the RNA was assessed electrophoretically.

Following preparation of the mRNAs from various tissues as described above, selection of mRNA with intact 5' ends and specific attachment of an oligonucleotide tag to the 5' end of such mRNA was performed using either a chemical or enzymatic approach. Both techniques take advantage of the presence of the "cap" structure, which characterizes the 5' end of most intact mRNAs and which comprises a guanosine generally methylated once, at the 7 position.

The chemical modification approach involves the optional elimination of the 2', 3'-cis diol of the 3' terminal ribose, the oxidation of the 2', 3'-cis diol of the ribose linked to the cap of the 5' ends of the mRNAs into a dialdehyde, and the coupling of the such obtained dialdehyde to a derivatized oligonucleotide tag. Further detail regarding the chemical approaches for obtaining mRNAs having intact 5' ends are disclosed in International Application No. WO96/34981 published November 7, 1996.

The enzymatic approach for ligating the oligonucleotide tag to the intact 5' ends of mRNAs involves the removal of the phosphate groups present on the 5' ends of uncapped incomplete mRNAs, the subsequent decapping of mRNAs having intact 5' ends and the ligation of the phosphate present at the 5' end of the decapped mRNA to an oligonucleotide tag. Further detail regarding the enzymatic approaches for obtaining mRNAs having intact 5' ends are disclosed in Dumas Milne Edwards J.B. (Doctoral Thesis of Paris VI University, Le clonage des ADNc complets: difficultés et perspectives nouvelles. Apports pour l'étude de la régulation de l'expression de la tryptophane hydroxylase de rat, 20 Dec. 1993), EP0 625572 and Kato *et al.*, *Gene* 150:243-250 (1994).

In both the chemical and the enzymatic approach, the oligonucleotide tag has a restriction enzyme site (e.g. an EcoRI site) therein to facilitate later cloning procedures.

Following attachment of the oligonucleotide tag to the mRNA, the integrity of the mRNA is examined by performing a Northern blot using a probe complementary to the oligonucleotide tag.

For the mRNAs joined to oligonucleotide tags using either the chemical or the enzymatic method, first strand cDNA synthesis is performed using an oligo-dT primer with reverse transcriptase. This oligo-dT primer can contain an internal tag of at least 4 nucleotides, which can be different from one mRNA preparation to another. Methylated dCTP is used for cDNA first strand synthesis to protect the internal EcoRI sites from digestion during subsequent steps. The first strand cDNA is precipitated using isopropanol after removal of RNA by alkaline hydrolysis to eliminate residual primers.

Second strand cDNA synthesis is conducted using a DNA polymerase, such as Klenow fragment and a primer corresponding to the 5' end of the ligated oligonucleotide. The primer is typically 20-25 bases in length. Methylated dCTP is used for second strand synthesis in order to protect internal EcoRI sites in the cDNA from digestion during the cloning process.

Following second strand synthesis, the full-length cDNAs are cloned into a phagemid vector, such as pBlueScript™ (Stratagene). The ends of the full-length cDNAs are blunted with T4 DNA polymerase (Biolabs) and the cDNA is digested with EcoRI. Since methylated dCTP is used during cDNA synthesis, the EcoRI site present in the tag is the only hemi-methylated site; hence the only site susceptible to EcoRI digestion. In some instances, to facilitate subcloning, an Hind III adapter is added to the 3' end of full-length cDNAs.

The full-length cDNAs are then size fractionated using either exclusion chromatography (AcA, Biosepra) or electrophoretic separation which yields 3 to 6 different fractions. The full-length cDNAs are then directionally cloned either into pBlueScript™ using either the EcoRI and SmaI restriction sites or, when the Hind III adapter is present in the full-length cDNAs, the EcoRI and Hind III restriction sites. The ligation mixture is transformed, preferably by electroporation, into bacteria, which are then propagated under appropriate antibiotic selection.

Clones containing the oligonucleotide tag attached to full-length cDNAs are selected as follows.

The plasmid cDNA libraries made as described above are purified (e.g. by a column available from Qiagen). A positive selection of the tagged clones is performed as follows. Briefly, in this selection procedure, the plasmid DNA is converted to single stranded DNA using phage F1 gene II endonuclease in combination with an exonuclease (Chang et al., *Gene* 127:95 (1993)) such as exonuclease III or T7 gene 6 exonuclease. The resulting single stranded DNA is then purified using paramagnetic beads as described by Fry et al., *Biotechniques* 13: 124 (1992). Here the single stranded DNA is hybridized with a biotinylated oligonucleotide having a sequence corresponding to the 3' end of the oligonucleotide tag. Preferably, the primer has a length of 20-25 bases. Clones including a sequence complementary to the biotinylated

oligonucleotide are selected by incubation with streptavidin coated magnetic beads followed by magnetic capture. After capture of the positive clones, the plasmid DNA is released from the magnetic beads and converted into double stranded DNA using a DNA polymerase such as ThermoSequenase™ (obtained from Amersham Pharmacia Biotech). Alternatively, protocols
5 such as the Gene Trapper™ kit (Gibco BRL) can be used. The double stranded DNA is then transformed, preferably by electroporation, into bacteria. The percentage of positive clones having the 5' tag oligonucleotide is typically estimated to be between 90 and 98% from dot blot analysis.

Following transformation, the libraries are ordered in microtiter plates and sequenced.
10 The *Arabidopsis* library was deposited at the American Type Culture Collection on January 7, 2000 as *E-coli* liba 010600" under the accession number PTA-1161.

EXAMPLE 2: SOUTHERN HYBRIDIZATIONS

The SDFs of the invention can be used in Southern hybridizations as described above. The following describes extraction of DNA from nuclei of plant cells, digestion of the
15 nuclear DNA and separation by length, transfer of the separated fragments to membranes, preparation of probes for hybridization, hybridization and detection of the hybridized probe.

The procedures described herein can be used to isolate related polynucleotides or for diagnostic purposes. Moderate stringency hybridization conditions, as defined above, are described in the present example. These conditions result in detection of hybridization
20 between sequences having at least 70% sequence identity. As described above, the hybridization and wash conditions can be changed to reflect the desired percentatge of sequence identity between probe and target sequences that can be detected.

In the following procedure, a probe for hybridization is produced from two PCR reactions using two primers from genomic sequence of *Arabidopsis thaliana*. As described
25 above, the particular template for generating the probe can be any desired template.

The first PCR product is assessed to validate the size of the primer to assure it is of the expected size. Then the product of the first PCR is used as a template, with the same pair of primers used in the first PCR, in a second PCR that produces a labeled product used as the probe.

30 Fragments detected by hybridization, or other bands of interest, can be isolated from gels used to separate genomic DNA fragments by known methods for further purification and/or characterization.

Buffers for nuclear DNA extraction

1. 10X HB

	1000 ml	
40 mM spermidine	10.2 g	Spermine (Sigma S-2876) and spermidine (Sigma S-2501)
10 mM spermine	3.5 g	Stabilize chromatin and the nuclear membrane
0.1 M EDTA (disodium)	37.2 g	EDTA inhibits nuclease
0.1 M Tris	12.1 g	Buffer
0.8 M KCl	59.6 g	Adjusts ionic strength for stability of nuclei

Adjust pH to 9.5 with 10 N NaOH. It appears that there is a nuclease present in leaves. Use of pH 9.5 appears to inactivate this nuclease.

2. 2 M sucrose (684 g per 1000 ml)

Heat about half the final volume of water to about 50°C. Add the sucrose slowly then bring the mixture to close to final volume; stir constantly until it has dissolved. Bring the solution to volume.

3. Sarkosyl solution (lyses nuclear membranes)

	<u>1000 ml</u>
N-lauroyl sarcosine (Sarkosyl)	20.0 g
0.1 M Tris	12.1 g
0.04 M EDTA (Disodium)	14.9 g

Adjust the pH to 9.5 after all the components are dissolved and bring up to the proper volume.

4. 20% Triton X-100
 80 ml Triton X-100
 320 ml 1xHB (w/o β -ME and PMSF)
 Prepare in advance; Triton takes some time to dissolve

5 A. Procedure

1. Prepare 1X H⁺ buffer (keep ice-cold during use)

	<u>1000 ml</u>
10X HB	100 ml
2 M sucrose	250 ml a non-ionic osmoticum
Water	634 ml

Added just before use:

100 mM PMSF*	10 ml a protease inhibitor; protects nuclear membrane proteins
β -mercaptoethanol	1 ml inactivates nuclease by reducing disulfide bonds

***100 mM PMSF**

(phenyl methyl sulfonyl fluoride, Sigma P-7626)

(add 0.0875 g to 5 ml 100% ethanol)

2. Homogenize the tissue in a blender (use 300-400 ml of 1xHB per blender). Be sure that you use 5-10 ml of HB buffer per gram of tissue. Blenders generate heat so be sure to keep the homogenate cold. It is necessary to put the blenders in ice periodically.
3. Add the 20% Triton X-100 (25 ml per liter of homogenate) and gently stir on ice for 20 min. This lyses plastid, but not nuclear, membranes.

6. Discard the dark green supernatant. The pellet will have several layers to it. One is starch; it is white and gritty. The nuclei are gray and soft. In the early steps, there may be a dark green and somewhat viscous layer of chloroplasts.

Wash the pellets in about 25 ml cold H buffer (with Triton X-100) and resuspend by swirling gently and pipetting. After the pellets are resuspended.

Pellet the nuclei again at 1200 - 1300 x g. Discard the supernatant.

Repeat the wash 3-4 times until the supernatant has changed from a dark green to a pale green. This usually happens after 3 or 4 resuspensions. At this point, the pellet is typically grayish white and very slippery. The Triton X-100 in these repeated steps helps to destroy the chloroplasts and mitochondria that contaminate the prep.

Resuspend the nuclei for a final time in a total of 15 ml of H buffer and transfer the suspension to a sterile 125 ml Erlenmeyer flask.

7. Add 15 ml, dropwise, cold 2% Sarkosyl, 0.1 M Tris, 0.04 M EDTA solution (pH 9.5) while swirling gently. This lyses the nuclei. The solution will become very viscous.
8. Add 30 grams of CsCl and gently swirl at room temperature until the CsCl is in solution. The mixture will be gray, white and viscous.
9. Centrifuge the solution at 11,400 x g at 4°C for at least 30 min. The longer this spin is, the firmer the protein pellicle.

10. The result is typically a clear green supernatant over a white pellet, and (perhaps) under a protein pellicle. Carefully remove the solution under the protein pellicle and above the pellet. Determine the density of the solution by weighing 1 ml of solution and add CsCl if necessary to bring to 1.57 g/ml. The solution contains dissolved solids (sucrose etc) and the refractive index alone will not be an accurate guide to CsCl concentration.
11. Add 20 μ l of 10 mg/ml EtBr per ml of solution.
12. Centrifuge at 184,000 x g for 16 to 20 hours in a fixed-angle rotor.
13. Remove the dark red supernatant that is at the top of the tube with a plastic transfer pipette and discard. Carefully remove the DNA band with another transfer pipette. The DNA band is usually visible in room light; otherwise, use a long wave UV light to locate the band.
14. Extract the ethidium bromide with isopropanol saturated with water and salt. Once the solution is clear, extract at least two more times to ensure that all of the EtBr is gone. Be very gentle, as it is very easy to shear the DNA at this step. This extraction may take a while because the DNA solution tends to be very viscous. If the solution is too viscous, dilute it with TE.
15. Dialyze the DNA for at least two days against several changes (at least three times) of TE (10 mM Tris, 1mM EDTA, pH 8) to remove the cesium chloride.
16. Remove the dialyzed DNA from the tubing. If the dialyzed DNA solution contains a lot of debris, centrifuge the DNA solution at least at 2500 x g for 10 min. and carefully transfer the clear supernatant to a new tube. Read the A260 concentration of the DNA.
17. Assess the quality of the DNA by agarose gel electrophoresis (1% agarose gel) of the DNA. Load 50 ng and 100 ng (based on the OD reading) and compare it with known

and good quality DNA. Undigested lambda DNA and a lambda-HindIII-digested DNA are good molecular weight makers.

Protocol for Digestion of Genomic DNA

Protocol:

1. The relative amounts of DNA for different crop plants that provide approximately a balanced number of genome equivalent is given in Table 3. Note that due to the size of the wheat genome, wheat DNA will be underrepresented. Lambda DNA provides a useful control for complete digestion.
2. Precipitate the DNA by adding 3 volumes of 100% ethanol. Incubate at -20°C for at least two hours. Yeast DNA can be purchased and made up at the necessary concentration, therefore no precipitation is necessary for yeast DNA.
3. Centrifuge the solution at $11,400 \times g$ for 20 min. Decant the ethanol carefully (be careful not to disturb the pellet). Be sure that the residual ethanol is completely removed either by vacuum desiccation or by carefully wiping the sides of the tubes with a clean tissue.
4. Resuspend the pellet in an appropriate volume of water. Be sure the pellet is fully resuspended before proceeding to the next step. This may take about 30 min.
5. Add the appropriate volume of 10X reaction buffer provided by the manufacturer of the restriction enzyme to the resuspended DNA followed by the appropriate volume of enzymes. Be sure to mix it properly by slowly swirling the tubes.
6. Set-up the lambda digestion-control for each DNA that you are digesting.
7. Incubate both the experimental and lambda digests overnight at 37°C . Spin down condensation in a microfuge before proceeding.
8. After digestion, add $2 \mu\text{l}$ of loading dye (typically 0.25% bromophenol blue, 0.25% xylene cyanol in 15% Ficoll or 30% glycerol) to the lambda-control digests and load

in 1% TPE-agarose gel (TPE is 90 mM Tris-phosphate, 2 mM EDTA, pH 8). If the lambda DNA in the lambda control digests are completely digested, proceed with the precipitation of the genomic DNA in the digests.

9. Precipitate the digested DNA by adding 3 volumes of 100% ethanol and incubating in -20°C for at least 2 hours (preferably overnight).

EXCEPTION: *Arabidopsis* and yeast DNA are digested in an appropriate volume; they don't have to be precipitated.

10. Resuspend the DNA in an appropriate volume of TE (e.g., 22 µl x 50 blots = 1100 µl) and an appropriate volume of 10X loading dye (e.g., 2.4 µl x 50 blots = 120 µl). Be careful in pipetting the loading dye - it is viscous. Be sure you are pipetting the correct volume.

Table 3

Some guide points in digesting genomic DNA.

Species	Genome Size	Size Relative to Arabidopsis	Genome Equivalent to 2 µg Arabidopsis DNA	Amount of DNA per blot
Arabidopsis	120 Mb	1X	1X	2 µg
Brassica	1,100 Mb	9.2X	0.54X	10 µg
Corn	2,800 Mb	23.3X	0.43X	20 µg
Cotton	2,300 Mb	19.2X	0.52X	20 µg
Oat	11,300 Mb	94X	0.11X	20 µg
Rice	400 Mb	3.3X	0.75X	5 µg
Soybean	1,100 Mb	9.2X	0.54X	10 µg
Sugarbeet	758 Mb	6.3X	0.8X	10 µg
Sweetclover	1,100 Mb	9.2X	0.54X	10 µg
Wheat	16,000 Mb	133X	0.08X	20 µg

Yeast	15 Mb	0.12X	1X	0.25 µg
-------	-------	-------	----	---------

Protocol for Southern Blot Analysis

The digested DNA samples are electrophoresed in 1% agarose gels in 1x TPE buffer. Low voltage; overnight separations are preferred. The gels are stained with EtBr and photographed.

1. For blotting the gels, first incubate the gel in 0.25 N HCl (with gentle shaking) for about 15 min.
2. Then briefly rinse with water. The DNA is denatured by 2 incubations. Incubate (with shaking) in 0.5 M NaOH in 1.5 M NaCl for 15 min.
3. The gel is then briefly rinsed in water and neutralized by incubating twice (with shaking) in 1.5 M Tris pH 7.5 in 1.5 M NaCl for 15 min.
4. A nylon membrane is prepared by soaking it in water for at least 5 min, then in 6X SSC for at least 15 min. before use. (20x SSC is 175.3 g NaCl, 88.2 g sodium citrate per liter, adjusted to pH 7.0.)
5. The nylon membrane is placed on top of the gel and all bubbles in between are removed. The DNA is blotted from the gel to the membrane using an absorbent medium, such as paper toweling and 6x SCC buffer. After the transfer, the membrane may be lightly brushed with a gloved hand to remove any agarose sticking to the surface.
6. The DNA is then fixed to the membrane by UV crosslinking and baking at 80°C. The membrane is stored at 4°C until use.

B. Protocol for PCR Amplification of Genomic Fragments in Arabidopsis**Amplification procedures:**

1. Mix the following in a 0.20 ml PCR tube or 96-well PCR plate:

Volume	Stock	Final Amount or Conc.
0.5 μ l	\sim 10 ng/ μ l genomic DNA ¹	5 ng
2.5 μ l	10X PCR buffer	20 mM Tris, 50 mM KCl
0.75 μ l	50 mM MgCl ₂	1.5 mM
1 μ l	10 pmol/ μ l Primer 1 (Forward)	10 pmol
1 μ l	10 pmol/ μ l Primer 2 (Reverse)	10 pmol
0.5 μ l	5 mM dNTPs	0.1 mM
0.1 μ l	5 units/ μ l Platinum Taq™ (Life Technologies, Gaithersburg, MD) DNA Polymerase	1 units
(to 25 μ l)	Water	

2. The template DNA is amplified using a Perkin Elmer 9700 PCR machine:

1) 94°C for 10 min. followed by

2) 5 cycles:	3) 5 cycles:	4) 25 cycles:
94 °C - 30 sec 62 °C - 30 sec 72 °C - 3 min	94 °C - 30 sec 58 °C - 30 sec 72 °C - 3 min	94 °C - 30 sec 53 °C - 30 sec 72 °C - 3 min

¹ Arabidopsis DNA is used in the present experiment, but the procedure is a general one.

- 5) 72°C for 7 min. Then the reactions are stopped by chilling to 4°C.

The procedure can be adapted to a multi-well format if necessary.

Quantification and Dilution of PCR Products:

1. The product of the PCR is analyzed by electrophoresis in a 1% agarose gel. A linearized plasmid DNA can be used as a quantification standard (usually at 50, 100, 200, and 400 ng). These will be used as references to approximate the amount of PCR products. HindIII-digested Lambda DNA is useful as a molecular weight marker. The gel can be run fairly quickly; e.g., at 100 volts. The standard gel is examined to determine that the size of the PCR products is consistent with the expected size and if there are significant extra bands or smeary products in the PCR reactions.
2. The amounts of PCR products can be estimated on the basis of the plasmid standard.
3. For the small number of reactions that produce extraneous bands, a small amount of DNA from bands with the correct size can be isolated by dipping a sterile 10-μl tip into the band while viewing through a UV Transilluminator. The small amount of agarose gel (with the DNA fragment) is used in the labeling reaction.

C. Protocol for PCR-DIG-Labeling of DNA

Solutions:

Reagents in PCR reactions (diluted PCR products, 10X PCR Buffer, 50 mM MgCl₂, 5 U/μl Platinum Taq Polymerase, and the primers)

10X dNTP + DIG-11-dUTP [1:5]: (2 mM dATP, 2 mM dCTP, 2 mM dGTP, 1.65 mM dTTP, 0.35 mM DIG-11-dUTP)

10X dNTP + DIG-11-dUTP [1:10]: (2 mM dATP, 2 mM dCTP, 2 mM dGTP, 1.81 mM dTTP, 0.19 mM DIG-11-dUTP)

10X dNTP + DIG-11-dUTP [1:15]: (2 mM dATP, 2 mM dCTP, 2 mM dGTP, 1.875 mM dTTP, 0.125 mM DIG-11-dUTP)

TE buffer (10 mM Tris, 1 mM EDTA, pH 8)

Maleate buffer: In 700 ml of deionized distilled water, dissolve 11.61 g maleic acid and 8.77 g NaCl. Add NaOH to adjust the pH to 7.5. Bring the volume to 1 L. Stir for 15 min. and sterilize.

10% blocking solution: In 80 ml deionized distilled water, dissolve 1.16g maleic acid. Next, add NaOH to adjust the pH to 7.5. Add 10 g of the blocking reagent powder (Boehringer Mannheim, Indianapolis, IN, Cat. no. 1096176). Heat to 60°C while stirring to dissolve the powder. Adjust the volume to 100 ml with water. Stir and sterilize.

1% blocking solution: Dilute the 10% stock to 1% using the maleate buffer.

Buffer 3 (100 mM Tris, 100 mM NaCl, 50 mM MgCl₂, pH9.5). Prepared from autoclaved solutions of 1M Tris pH 9.5, 5 M NaCl, and 1 M MgCl₂ in autoclaved distilled water.

Procedure:

1. PCR reactions are performed in 25 μ l volumes containing:

PCR buffer	1X
MgCl ₂	1.5 mM
10X dNTP + DIG-11-dUTP	1X (please see the note below)
Platinum Taq™ Polymerase	1 unit
10 pg probe DNA	
10 pmol primer 1	

Note:Use for:

10X dNTP + DIG-11-dUTP (1:5)	< 1 kb
10X dNTP + DIG-11-dUTP (1:10)	1 kb to 1.8 kb
10X dNTP + DIG-11-dUTP (1:15)	> 1.8 kb

2. The PCR reaction uses the following amplification cycles:

- 1) 94°C for 10 min.

2) 5 cycles:	3) 5 cycles:	4) 25 cycles:
95°C - 30 sec 61°C - 1 min 73°C - 5 min	95°C - 30 sec 59°C - 1 min 75°C - 5 min	95°C - 30 sec 51°C - 1 min 73°C - 5 min

- 5) 72°C for 8 min. The reactions are terminated by chilling to 4°C (hold).

3. The products are analyzed by electrophoresis- in a 1% agarose gel, comparing to an aliquot of the unlabelled probe starting material.
4. The amount of DIG-labeled probe is determined as follows:

Make serial dilutions of the diluted control DNA in dilution buffer (TE: 10 mM Tris and 1 mM EDTA, pH 8) as shown in the following table:

DIG-labeled control DNA starting conc.	Stepwise Dilution	Final Conc. (Dilution Name)
5 ng/μl	1 μl in 49 μl TE	100 pg/μl (A)
100 pg/μl (A)	25 μl in 25 μl TE	50 pg/μl (B)
50 pg/μl (B)	25 μl in 25 μl TE	25 pg/μl (C)
25 pg/μl (C)	20 μl in 30 μl TE	10 pg/μl (D)

- a. Serial dilutions of a DIG-labeled standard DNA ranging from 100 pg to 10 pg are spotted onto a positively charged nylon membrane, marking the membrane lightly with a pencil to identify each dilution.
- b. Serial dilutions (e.g., 1:50, 1:2500, 1:10,000) of the newly labeled DNA probe are spotted.
- c. The membrane is fixed by UV crosslinking.
- d. The membrane is wetted with a small amount of maleate buffer and then incubated in 1% blocking solution for 15 min at room temp.
- e. The labeled DNA is then detected using alkaline phosphatase conjugated anti-DIG antibody (Boehringer Mannheim, Indianapolis, IN, cat. no. 1093274) and an NBT substrate according to the manufacture's instruction.
- f. Spot intensities of the control and experimental dilutions are then compared to estimate the concentration of the PCR-DIG-labeled probe.

D. Prehybridization and Hybridization of Southern Blots**Solutions:**

100% Formamide purchased from Gibco

20X SSC (1X = 0.15 M NaCl, 0.015 M Na₃citrate)

per L: 175 g NaCl

87.5 g Na₃citrate·2H₂O

20% Sarkosyl (N-lauroyl-sarcosine)

20% SDS (sodium dodecyl sulphate)

10% Blocking Reagent: In 80 ml deionized distilled water, dissolve 1.16 g maleic acid. Next, add NaOH to adjust the pH to 7.5. Add 10 g of the blocking reagent powder. Heat to 60°C while stirring to dissolve the powder. Adjust the volume to 100 ml with water. Stir and sterilize.

Prehybridization Mix:

Final Concentration	Components	Volume (per 100 ml)	Stock
50%	Formamide	50 ml	100%
5X	SSC	25 ml	20X
0.1%	Sarkosyl	0.5 ml	20%
0.02%	SDS	0.1 ml	20%
2%	Blocking Reagent	20 ml	10%
	Water	4.4 ml	

General Procedures:

- Place the blot in a heat-sealable plastic bag and add an appropriate volume of prehybridization solution (30 ml/100cm²) at room temperature. Seal the bag with a heat sealer, avoiding bubbles as much as possible. Lay down the bags in a large plastic tray (one tray can accommodate at least 4–5 bags). Ensure that the bags are

lying flat in the tray so that the prehybridization solution is evenly distributed throughout the bag. Incubate the blot for at least 2 hours with gentle agitation using a waver shaker.

2. Denature DIG-labeled DNA probe by incubating for 10 min. at 98°C using the PCR machine and immediately cool it to 4°C.

3. Add probe to prehybridization solution (25 ng/ml; 30 ml = 750 ng total probe) and mix well but avoid foaming. Bubbles may lead to background.

4. Pour off the prehybridization solution from the hybridization bags and add new prehybridization and probe solution mixture to the bags containing the membrane.

5. Incubate with gentle agitation for at least 16 hours.

6. Proceed to medium stringency post-hybridization wash:

Three times for 20 min. each with gentle agitation using 1X SSC, 1% SDS at 60°C.

All wash solutions must be prewarmed to 60°C. Use about 100 ml of wash solution per membrane.

To avoid background keep the membranes fully submerged to avoid drying in spots; agitate sufficiently to avoid having membranes stick to one another.

7. After the wash, proceed to immunological detection and CSPD development.

E. Procedure for Immunological Detection with CSPD

Solutions:

Buffer 1: Maleic acid buffer (0.1 M maleic acid, 0.15 M NaCl; adjusted to pH 7.5 with NaOH)

Washing buffer: Maleic acid buffer with 0.3% (v/v) Tween 20.

Blocking stock solution 10% blocking reagent in buffer 1. Dissolve (10X concentration): blocking reagent powder (Boehringer Mannheim, Indianapolis, IN, cat. no. 1096176) by constantly stirring on a 65°C heating block or heat in a microwave, autoclave and store at 4°C.

Buffer 2
(1X blocking solution): Dilute the stock solution 1:10 in Buffer 1.

Detection buffer: 0.1 M Tris, 0.1 M NaCl, pH 9.5

Procedure:

1. After the post-hybridization wash the blots are briefly rinsed (1-5 min.) in the maleate washing buffer with gentle shaking.
2. Then the membranes are incubated for 30 min. in Buffer 2 with gentle shaking.
3. Anti-DIG-AP conjugate (Boehringer Mannheim, Indianapolis, IN, cat. no. 1093274) at 75 mU/ml (1:10,000) in Buffer 2 is used for detection. 75 ml of solution can be used for 3 blots.
4. The membrane is incubated for 30 min. in the antibody solution with gentle shaking.
5. The membrane are washed twice in washing buffer with gentle shaking. About 250 mls is used per wash for 3 blots.
6. The blots are equilibrated for 2-5 min in 60 ml detection buffer.
7. Dilute CSPD (1:200) in detection buffer. (This can be prepared ahead of time and stored in the dark at 4°C).

The following steps must be done individually. Bags (one for detection and one for exposure) are generally cut and ready before doing the following steps.

8. The blot is carefully removed from the detection buffer and excess liquid removed without drying the membrane. The blot is immediately placed in a bag and 1.5 ml of CSPD solution is added. The CSPD solution can be spread over the membrane. Bubbles present at the edge and on the surface of the blot are typically removed by gentle rubbing. The membrane is incubated for 5 min. in CSPD solution.
9. Excess liquid is removed and the membrane is blotted briefly (DNA side up) on Whatman 3MM paper. Do not let the membrane dry completely.
10. Seal the damp membrane in a hybridization bag and incubate for 10 min at 37°C to enhance the luminescent reaction.
11. Expose for 2 hours at room temperature to X-ray film. Multiple exposures can be taken. Luminescence continues for at least 24 hours and signal intensity increases during the first hours.

Example 3: Transformation of Carrot Cells

Transformation of plant cells can be accomplished by a number of methods, as described above. Similarly, a number of plant genera can be regenerated from tissue culture following transformation. Transformation and regeneration of carrot cells as described herein is illustrative.

Single cell suspension cultures of carrot (*Daucus carota*) cells are established from hypocotyls of cultivar Early Nantes in B₅ growth medium (O.L. Gamborg et al., *Plant Physiol.* 45:372 (1970)) plus 2,4-D and 15 mM CaCl₂ (B₅-44 medium) by methods known in the art. The suspension cultures are subcultured by adding 10 ml of the suspension culture to 40 ml of B₅-44 medium in 250 ml flasks every 7 days and are maintained in a shaker at 150 rpm at 27 °C in the dark.

The suspension culture cells are transformed with exogenous DNA as described by Z. Chen et al. *Plant Mol. Bio.* 36:163 (1998). Briefly, 4-days post-subculture cells are incubated with cell wall digestion solution containing 0.4 M sorbitol, 2% driselase, 5mM MES (2-[N-Morpholino] ethanesulfonic acid) pH 5.0 for 5 hours. The digested cells are pelleted gently at 60 xg for 5 min. and washed twice in W5 solution containing 154 mM NaCl, 5 mM KCl, 125 mM CaCl₂ and 5mM glucose, pH 6.0. The protoplasts are suspended in MC solution

containing 5 mM MES, 20 mM CaCl₂, 0.5 M mannitol, pH 5.7 and the protoplast density is adjusted to about 4×10^6 protoplasts per ml.

15-60 µg of plasmid DNA is mixed with 0.9 ml of protoplasts. The resulting suspension is mixed with 40% polyethylene glycol (MW 8000, PEG 8000), by gentle inversion a few times at room temperature for 5 to 25 min. Protoplast culture medium known in the art is added into the PEG-DNA-protoplast mixture. Protoplasts are incubated in the culture medium for 24 hour to 5 days and cell extracts can be used for assay of transient expression of the introduced gene. Alternatively, transformed cells can be used to produce transgenic callus, which in turn can be used to produce transgenic plants, by methods known in the art. See, for example, Nomura and Komamine, *Plt. Phys.* 79:988-991 (1985), *Identification and Isolation of Single Cells that Produce Somatic Embryos in Carrot Suspension Cultures*.

An additional deposit of an *E. coli* Library, *E. coli*LibA021800, was made at the American Type Culture Collection in Manassas, Virginia, USA on February 22, 2000 to meet the requirements of Budapest Treaty for the international recognition of the deposit of microorganisms.

The invention being thus described, it will be apparent to one of ordinary skill in the art that various modifications of the materials and methods for practicing the invention can be made. Such modifications are to be considered within the scope of the invention as defined by the following claims.

Each of the references from the patent and periodical literature cited herein is hereby expressly incorporated in its entirety by such citation.

0966980-101300